An Outcome Screening Procedure for Small Areas

Pavlina Rumcheva
Donald Malec, Peter Meyer, Nathaniel Schenker

National Center for Health Statistics
Centers for Disease Control and Prevention

Federal Committee on Statistical Methodology Policy Seminar
Washington, D.C., December 4, 2012

The findings and conclusions in this presentation are those of the authors and do not necessarily represent the views of the National Center for Health Statistics, CDC.

Outline

- **Motivation and Goal**
- **Data and Outcomes**
- **Methods and Results**
- **Conclusions and Future Work**

Motivation and Goal

- estimates of population characteristics are desired for geographic regions and population subgroups for which sufficient data may not be available to obtain reliable direct estimates
- model-based estimates can be produced for small areas defined by location and demographic factors using auxiliary data
- develop a procedure to determine outcomes with large variability across small areas
- allocate resources for the development and estimation of small-area statistical models

Data and Outcomes

The National Health Interview Survey (2010)

Some outcomes of interest:

- uninsured at time of interview
- uninsured but had insurance within the past year
- has private insurance
- has flexible spending account
- enrolled in a high-deductible health plan
- medical care delayed in the past year due to cost
- poor health
- limited in any way

Data and Outcomes

Outcomes of interest are binary.

Example: Has a flexible spending account? Yes/No.

National estimates for proportions are generally available.

Estimates are desired for small areas determined by:

- geographic regions: State, metropolitan area, county.
- demographic factors: race, gender, age.

The goal is to identify outcomes with large small-area variability. This information will be used to allocate resources for small-area model development and estimation.

Methods and Results

Denote $p_1, \ldots, p_n$ proportions for a single outcome in n small areas.

What is the amount of variability among the proportions?

Coffey, Feingold and Bromberg(1988) developed descriptive measure of disparity among proportions relative to maximum possible disparity:

$$V = \frac{1}{n}\sum_{i=1}^{n}(p_i - \bar{p})^2, \quad \bar{p} = \frac{1}{n}\sum_{i=1}^{n} p_i,$$

$$\mathbf{RELV} = \sqrt{\frac{\mathbf{V}}{\mathbf{max(V|\bar{p})}}}, \quad \textbf{Relative Variability},$$

$$\max(V|\bar{p}) = \bar{p}(1-\bar{p}) - \frac{1}{n}r(1-r), \quad r = n\bar{p} - [n\bar{p}].$$

Methods and Results

$$\textbf{Relative Variability}: \quad \textbf{RELV} = \sqrt{\frac{\mathbf{V}}{\mathbf{max}(\mathbf{V}|\bar{\mathbf{p}})}}$$

- Maximum possible variability depends upon $\bar{p}$
- $0 \leq RELV \leq 1$
- *RELV* can achieve its minimum and maximum values 0 and 1 for all combinations of n and $\bar{p}$
- *RELV* indicates position of $p_1, \ldots, p_n$ relative to extremes of no variability (all proportions equal) or maximum variability (proportions as varied as possible given $\bar{p}$)

Methods and Results

Small area: State i, $i = 1, \ldots, n$.

Observation: Outcome j, Proportion p_{ij}, $i = 1, \ldots, n; j = 1, \ldots, k$.

Denote $\underset{\sim}{p_i} = \{p_{i1}, \ldots, p_{ik}\}$ unknown population proportions in State i

$\underset{\sim}{\widehat{p}_i} = \{\widehat{p}_{i1}, \ldots, \widehat{p}_{ik}\}$ design-based proportion estimates.

Consider the model:

$$\underset{\sim}{\widehat{p}_i} \overset{ind}{\sim} N\left(\underset{\sim}{p_i}, \widehat{\Sigma}_i\right)$$

$$p_{ij} \overset{iid}{\sim} Uniform(0, 1)$$

$$i = 1, \ldots, n;\ j = 1, \ldots, k$$

where $\widehat{\Sigma}_i$ is $k \times k$ design-based covariance matrix for State i.

Then the posterior is $\underset{\sim}{p_i}|\underset{\sim}{\widehat{p}_i} \overset{ind}{\sim} N\left(\underset{\sim}{\widehat{p}_i}, \widehat{\Sigma}_i\right)$ truncated (0,1).

Methods and Results

Denote $RELV_j$ relative variability of Outcome j across States.

Obtain posterior mean and standard error of $RELV_j$.

Rank variability of Outcomes across States according to $RELV_j$.

Determine posterior probabilities of rankings.

Outcomes with large $RELV_j$ are candidates for small-area modeling.

Outcomes with small $RELV_j$ can be reported with national rates.

Methods and Results

- use R software and Survey package to obtain design-based estimates and conduct posterior inference
- remove variability due to demographic factors such as age, gender, race, using synthetic estimates based on national rates
- shift proportion estimates of 0 or 1, and determine conservative standard errors based on Wilson (1927) confidence interval
- transform proportion using $arcsin(\sqrt{\hat{p}})$ to stabilize variance
- determine variance of transformed proportion using Taylor series
- use rejection sampling to generate draws from posterior of transformed proportion

Methods and Results

If proportion estimate equals 0 or 1: shift estimate toward center of the interval (0,1) and obtain conservative estimate of standard error.

Wilson (1927) Confidence Interval:

Invert acceptance region of large-sample two-tailed test for $H_0 : p = p_0$ using null standard error to obtain $100\alpha\%$ CI:

$$\frac{\hat{p} + z_\alpha^2/2s}{1 + z_\alpha^2/s} \pm \frac{\sqrt{\hat{p}(1-\hat{p})z_\alpha^2/s + z_\alpha^4/4s^2}}{1 + z_\alpha^2/s}$$

where $\hat{p}$-observed proportion, s-sample size, z_α-standard normal quantile. Examples:

$\hat{p} = 0, s = 10 \Rightarrow 95\%\text{CI is } (0, 0.28),\ \hat{p}_{new} = 0.14,\ \text{var}(\hat{p}_{new}) = 0.07^2$

$\hat{p} = 1, s = 20 \Rightarrow 95\%\text{CI is } (0.84, 1),\ \hat{p}_{new} = 0.92,\ \text{var}(\hat{p}_{new}) = 0.04^2$

Methods and Results

Relative Variability Posterior Mean and Standard Deviation

Outcome	RELV %	StDev %
High Deductible Health Plan with HSA	28.5	1.8
Uninsured, Had Insurance more than 1 Year Ago	24.5	2.2
High Deductible Health Plan	24.2	2.6
Uninsured, Had Insurance in Past Year	23.7	2.8
Private Insurance	18.2	1.3
Flexible Spending Account	17.5	1.3
Uninsured at Time of Interview	16.5	1.0
Fair/Poor Health	11.4	0.8
No Medical Care in Past Year due to Cost	11.3	1.2
Medical Care Delayed in Past Year due to Cost	11.1	1.0
Limited in any Way	11.1	0.9
More than 10 Doctor Visits	9.9	1.2

Methods and Results

Adjusted Relative Variability RELV-a and Its Ratio with RELV

Outcome	RELV-a %	Ratio %
High Deductible Health Plan with HSA	22.6	79
Uninsured, Had Insurance in Past Year	20.1	85
High Deductible Health Plan	19.8	82
Uninsured, Had Insurance more than 1 Year Ago	18.4	75
Flexible Spending Account	15.2	87
Uninsured at Time of Interview	13.3	81
Private Insurance	10.7	59
No Medical Care in Past Year due to Cost	10.7	95
Medical Care Delayed in Past Year due to Cost	10.4	93
Fair/Poor Health	10.0	87
Limited in any Way	9.6	87
More than 10 Doctor Visits	9.0	92

Methods and Results

Relative Variability Posterior Mean and St. Dev. for Asians

	Outcome	RELV %	StDev %
1	High Deductible Health Plan with HSA	47.8	3.4
2	Uninsured, Had Insurance in Past Year	43.6	3.8
3	More than 10 Doctor Visits	41.9	5.1
4	Uninsured, Had Insurance more 1 Year Ago	41.6	3.3
5	Limited in any Way	41.0	4.5
6	High Deductible Health Plan	40.6	3.2
7	Medical Care Delayed in Past Year due to Cost	40.5	4.5
8	Flexible Spending Account	40.4	3.7
9	No Medical Care in Past Year due to Cost	40.1	4.6
10	Fair/Poor Health	37.1	4.4
11	Private Insurance	34.3	3.0
12	Uninsured at Time of Interview	32.5	3.2

Methods and Results

Posterior Order of Outcome Variability for Asians

P(RELV Outcome in Row $>$ RELV Outcome in Column)

	O2	O3	O4	O5	O6	O7	O8	O9	O10	O11	O12
O1	0.79	0.84	0.90	0.89	0.95	0.91	0.94	0.91	0.97	1.00	1.00
O2		0.61	0.67	0.66	0.73	0.70	0.73	0.73	0.87	0.97	0.99
O3			0.53	0.55	0.58	0.60	0.59	0.61	0.77	0.90	0.94
O4				0.54	0.58	0.58	0.60	0.60	0.78	0.95	0.98
O5					0.53	0.54	0.54	0.55	0.73	0.90	0.93
O6						0.51	0.50	0.53	0.74	0.92	0.96
O7							0.49	0.52	0.70	0.87	0.92
O8								0.52	0.72	0.90	0.94
O9									0.67	0.86	0.91
O10										0.70	0.81
O11											0.68

Conclusions and Future Work

Conclusions:

- stable estimates for State relative variability based on posterior
- quick screening: identify outcomes with large relative variability ignoring correlation
- cost-cutting screening: use a covariance model for outcomes with large variability for better accuracy
- determine outcome relative variability by domain
- covariate selection: examine outcome variability adjusting for covariates as a quick way to screen variables for statistical model

Future Work:

- evaluate more outcomes
- validate normality assumption
- examine other measures of variability