Using BRFSS Data to Estimate County-level Colorectal Cancer Screening Prevalence in Missouri

J Du, MA1,2, CL Schmalz, PhD1,1; J Jackson-Thompson, MSPH, PhD1,3,4

1Missouri Cancer Registry and Research Center (MCR-ARC); 2University of Missouri-Columbia (MU), College of Arts & Sciences, Dept. of Statistics; 3MU School of Medicine, Dept. of Health Management & Informatics; 4MU Informatics Institute, Columbia, Missouri

BACKGROUND

In the US, colorectal cancer (CRC) is the 3rd most common cancer in both men and women.

- Colorectal cancer screening (CRCS) is recommended for people over 50 years of age.
- Behavioral Risk Factor Surveillance System (BRFSS) data in 2012 show that 66.5% of people in Missouri aged 50 and older have had the screening. However, county-level CRCS prevalence cannot be directly obtained from BRFSS due to small or even zero sample sizes.

METHODS: NOTATIONS

Respondents were classified into 12 groups based on age (50–64, 65–74, 75+), gender and race (white, non-white). Let

- \( n_{ijkl} \) : sample size (number of respondents) in county \( i \) in \( j \) age group \( j \) \( i \) \( j \) \( k \) \( l \)
- \( N_{ijkl} \) : true population size for county \( i \) \( j \) \( j \) \( k \) \( l \) which is obtained from Census data;
- \( y_{ijkl} \) : number of respondents who have had CRCs for category \( i j k l \);
- \( Y_{ijkl} \) : true total population for who have had CRCs for category \( i j k l \);
- \( p_{ijkl} \) : proportion of people who have had CRCs for category \( i j k l \).

METHODS: MODELS

A Bayesian binomial regression framework to estimate \( Y_{ijkl} \):

\[
Y_{ijkl} \sim \text{Binomial}(n_{ijkl}; p_{ijkl})
\]

\[
\logit(p_{ijkl}) = \alpha_{ij} + \beta_{ijkl} + \theta_i + \psi
\]

where \( \epsilon_{ijkl} \sim \text{Normal}(0, \delta_i) \) is the error term and

- \( \alpha_{ij} \) is the intercept for region \( i \) where county \( i \) belongs to;
- \( \beta = (\beta_1, \beta_2, \beta_3) \) are the age effects;
- \( \gamma = (\gamma_1, \gamma_2) \) are the gender effects;
- \( \theta = (\theta_1, \theta_2) \) are the race effects;
- \( \psi \) contains some county attribute effects like medium income, percentage of people below high school, etc.

METHODS: ESTIMATION

A Markov Chain Monte Carlo algorithm was used to obtain posterior samples for \( p_{ijkl} \).

METHODS: RESULTS

The county-level CRCS prevalence estimates from BRFSS generally agree with those from CLS, with an average 0.50%-point difference across all counties in MO.

METHODS: CONCLUSIONS

- The differences between BRFSS and CLS for counties with high/low CRCS prevalence are still noticeably large.
- In BRFSS, small or zero sample sizes for counties in Missouri potentially produce biased estimates. It is hard to estimate a whole county’s prevalence based only on several or tens of people.
- When BRFSS is the only source to estimate county-level prevalence, our model can still provide reasonable estimates at county level.
- We also used models in Cadwell, et al. (2010) to obtain CRCS prevalence estimates. However, due to small sample sizes in Missouri compared to all samples in US, covariances among groups of people were hard to estimate, which added more uncertainty compared to our model.
- We classified people into 12 groups in our analysis. However, when detailed population sizes are available, finer classification with more demographic variables may help improve the results.
- In our evaluation of the results, we treat CLS (2011) as the true prevalence for comparison. However, the uncertainty from CLS itself was not considered.

CONTACT: J Du (jdx6@email.missouri.edu)