Small Area Estimation Projects Are Definitely Not Small!

Yulei He Beresovu
National Center for Health Statistics
Centers for Disease Control and Prevention

Acknowledgements: Makram Talih, Don Malec, and Speakers/Authors of this session

*Disclaimer: The findings and conclusions in this study are those of the authors and do not necessarily represent the views of the Centers for Disease Control and Prevention.

Outline

- A Brief Review
- Main Notes
- Conclusion

A Brief Review

- Chakraborty et al.: Small domain estimation using probability and non-probability survey data
  - **Non-probability**
- Erciulescu et al.: On increasing the number of county-level crop estimates
  - **Increasing the number**
- Gershunskaya et al.: Small area co-modeling of point estimates and their variances for domains
  - **Co-modeling**
- Liu et al.: Small area estimation for measures related to tobacco use and policies using the tobacco use supplement to the current population survey
  - **Measures**

Key Components of SAE

- Data
- Modeling
- Estimation
- Diagnostics and Processing
- Publication

Data

- A single data source or multiple data sources?
  - All papers seem to use information from multiple data sources
- How to use information from multiple data sources?
  - The main outcome(s) of interests are collected from a probability survey
  - All papers used information from administrative sources (non-probability sample) as auxiliary variables (covariates)
  - Chakraborty et al. also proposed to model the same outcome (e.g., food allergy) from both data sources jointly
    - Assume that estimate from the probability is unbiased and that from the non-probability sample can be biased
    - How does this compare with the method that treats the non-probability estimate as a covariate (e.g., Erciulescu et al.)?

Modeling

- All papers used the Fay-Herriot (FH) modeling framework (i.e., area-level model)
  - How about unit-level models?
- All papers applied FH models to a single outcome or applied to multiple outcomes individually
  - How about multivariate FH models to account for the correlations among multiple measures (e.g., Liu et al.)?
- Can we sometime consider spatial correlations among the neighboring areas in addition to the typical between-area random effects (e.g., Erciulescu et al.)?

Modeling, ctd.

- A typical FH model

$$\hat{Y}_i|\theta_i \sim N(\theta_i, \sigma_i^2) \quad (1)$$

$$\theta_i|\beta, \tau^2 \sim N(X_i\beta, \tau^2) \quad (2)$$

- Often assume σ_i^2 is known and set it to a design-based estimate $\hat{\sigma}_i^2$
  - $\hat{\sigma}_i^2$ can be very unreliable with small samples at area i
  - Liu et al. proposed a practical solution
- Gershunskaya et al. proposed a sophisticated strategy by co-modeling the point and variance estimates, and they also relaxed the normality assumption of the 2nd-stage modeling
- Incomplete ($\hat{Y}_i$, $\hat{\sigma}_i^2$) or X_i
  - Erciulescu et al. proposed a practical solution
  - Missing data imputation?
- Selection of covariates X_i

Estimation

- All papers used the Bayesian estimation for their models
- What software packages?
  - Gershunskaya et al. used R Stan
  - WinBUGS? INLA? SAS PROC MCMC? Or authors' own code?
- Can authors show some sample code in their future publications?

Diagnostics and Processing

- All papers showed that model-based estimates are more precise than direct estimates
  - Which part of the improvement is attributable to the borrowing information from other areas, and which part is attributable to the use of auxiliary variables?
- Advanced Bayesian diagnostics (e.g., Liu et al. and Gershunskaya et al.)
- Benchmarking is a common exercise in SAE (e.g., Erciulescu et al.)

Publication

- One picture (map) is worth a thousand words (Erciulescu et al. and Liu et al.)
- Non-statistical challenges in the process of releasing small-area estimates to the public (Liu et al.)
  - Logistic issues?
  - Documenting and maintaining data
  - Disclosure risk?

Conclusion

- Fantastic work from all papers
- A great learning experience for me
- SAE is a very important, interesting, and challenging topic area
- More relevant work need to be and can be done