

# Applications of a Small Area Model for a Voting Rights Act Tabulation

Patrick M. Joyce, Donald Malec<sup>†</sup>, Roderick J. A. Little,  
Aaron Gilary

Center for Statistical Research and Methodology  
U.S. Census Bureau

<sup>†</sup>National Center for Health Statistics



## Disclaimers

### United States Census Bureau Disclaimer

This report is released to inform interested parties of research and to encourage discussion. The views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

### National Center for Health Statistics Disclaimer

The findings and conclusions in this paper are those of the authors and do not necessarily represent the official position of the National Center for Health Statistics, Centers for Disease Control and Prevention.

## Section 203 of the Voting Rights Act

### What is Section 203?

- ▶ Requires, when statutory conditions exist, jurisdictions to provide election materials in languages other than English.
- ▶ [Statutory] Target: Hispanic, Asian, and American Indian Alaska Native voting age citizens.
- ▶ Responsibility lies with the Director of the Census Bureau to determine if these conditions exist.
- ▶ Notice of determination is issued via Federal Register.

## Section 203 of the Voting Rights Act

Federal Register Notice: October 11, 2011

Section 203 mandates that a state or political subdivision must provide language assistance to voters if more than **five (5) percent** of voting age citizens are members of a **single** language minority group and **do not** “speak or understand English adequately enough to participate in the electoral process” and if the **rate** of those citizens who have **not completed the fifth grade** is higher than the national rate of voting age citizens who have not completed the fifth grade.

## Section 203 of the Voting Rights Act

Additional coverage conditions:

- ▶ Jurisdictions where the number of voting age limited-English proficient language minority group citizens exceeds 10,000.
- ▶ State coverage.
- ▶ American Indian Areas and Alaska Native Areas.

## Targeted Geographies

- ▶ 50 States.
- ▶ 7892 Jurisdictions...
  - ▶ 5082 MCDs in CT, ME, MA, MI, NH, RI, WI, and VT.
  - ▶ 2809 counties in remaining 42 states.
  - ▶ 1 District of Columbia.
- ▶ 919 American Indian Area and Alaska Native Areas.

## Previous Application of Section 203

- ▶ Previous determinations under Section 203 were made after the Census.
- ▶ Questions relating to citizenship, limited-English proficiency, and low education (illiteracy) were addressed by tabulation from the Census Long Form.
- ▶ Law was adjusted to reflect change of instrumentation to the American Community Survey (ACS).
- ▶ Original plan was to make determinations using direct survey-weighted estimates.

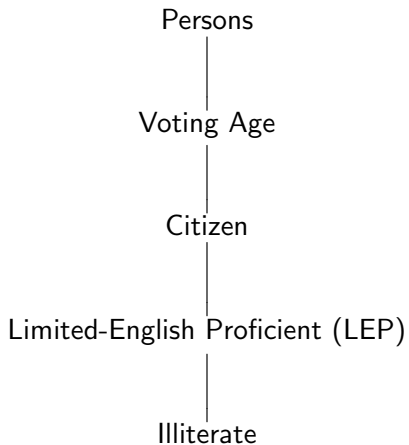
## Director's Challenge to Statistical Researchers

Challenge: Develop a small area model to provide targeted answers to statutory questions.

- ▶ Statutory questions can be seen in stages.



## Director's Challenge to Statistical Researchers



## Director's Challenge to Statistical Researchers

Challenge: Develop a small area model to provide targeted answers to statutory questions.

- ▶ Statutory questions can be seen in stages.
- ▶ Traditional small area models (Fay and Herriot, 1979) are not appropriate for binary outcomes.
- ▶ Race and ethnicity, being of interest, plays a large role in model development.

## Available Data Sources

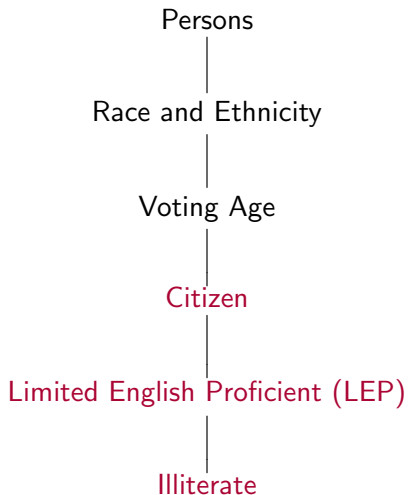
- ▶ American Community Survey (ACS)
- ▶ Decennial (2010) Census

	ACS 5-year ('05-'09)	2010 Census
Age	Yes	Yes
Race/Ethnicity	Yes	Yes
Citizenship	Yes	No
Limited-English	Yes	No
Illiteracy	Yes	No

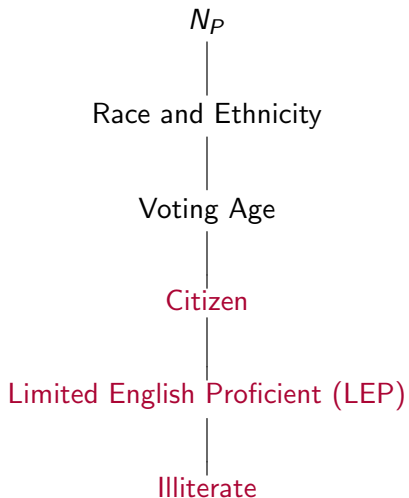
## Portents of a Model

- ▶ Determinations require tabulations of counts by age and race/ethnicity.
- ▶ Census provides population-based controls on race/ethnicity and age.
- ▶ Remaining unknowns must be addressed by ACS.
- ▶ **Provide a small area model, at the jurisdiction level, for each stage of citizenship, limited-English proficiency, and illiteracy by race/ethnicity.**

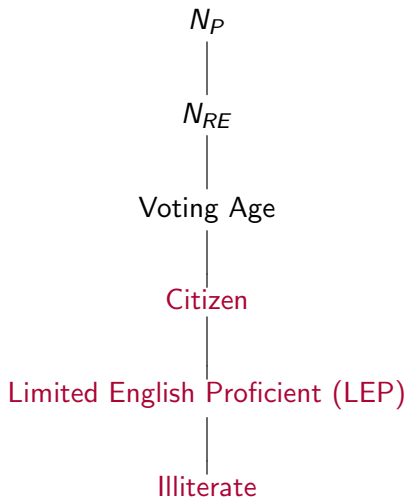
## Model Structure



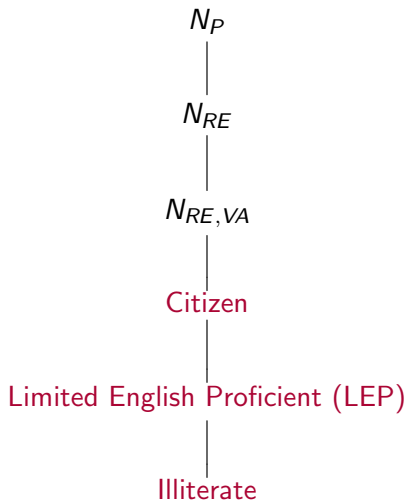
## Model Structure



## Model Structure

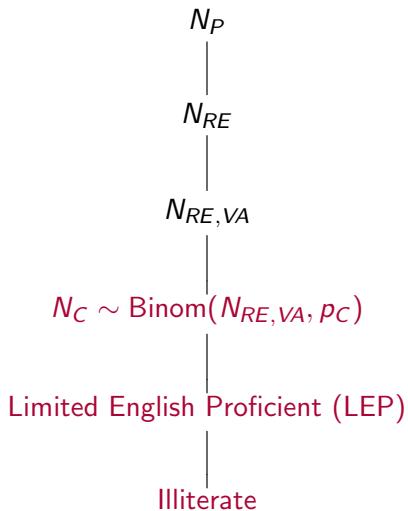


## Model Structure

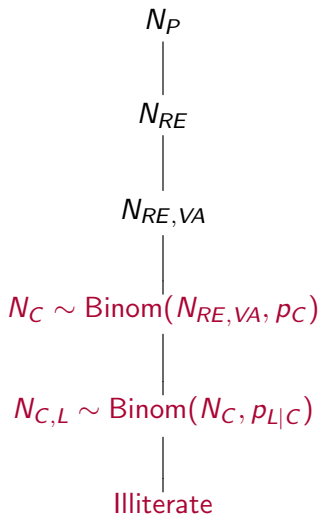




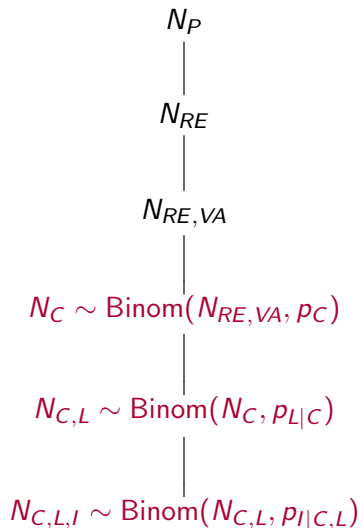
## Model Structure



## Model Structure



## Model Structure



## Small Area Estimation: The Plan

- ▶ Model is adjusted for design properties.
- ▶ Use beta distribution for random-effect (small area) term(s). This provides estimation shrinkage.
- ▶ Further shrinkage provided by placing jurisdictions into separate classes and using a separate small area model for each class.
- ▶ A model is fitted to ACS- and Census-based covariates to create predicted proportions for jurisdictions.
- ▶ Point estimates follow from appropriate multiplication and tabulation.

## Design-Effect and Likelihood

- ▶ Design-effect adjustment necessary to relate design features into the model.
- ▶ A design-effect is calculated at the national level and applied to form a jurisdiction level effective sample size.
- ▶ Jurisdictional effective sample size is combined with the design estimate to form effective response.

## Beta Hierarchy

- ▶ For each stage and for each model place a Beta distribution on the parameter  $p$  across the jurisdictions. This distribution has hyper-parameters  $\alpha$  and  $\beta$ .
- ▶ Hyper-parameters are estimated by finding the marginal distribution w.r.t. the data,  $\alpha$ , and  $\beta$  and then taking the maximum likelihood.
- ▶ This results in estimates of the form:

$$\hat{p} = \frac{x + \hat{\alpha}}{n + \hat{\alpha} + \hat{\beta}}.$$

## Beta Hierarchy

- ▶ For each stage and for each model place a Beta distribution on the parameter  $p$  across the jurisdictions. This distribution has hyper-parameters  $\alpha$  and  $\beta$ .
- ▶ Hyper-parameters are estimated by finding the marginal distribution w.r.t. the data,  $\alpha$ , and  $\beta$  and then taking the maximum likelihood.
- ▶ This results in estimates of the form:

$$\hat{p} = \frac{x}{n} \left( \frac{n}{n + \hat{\alpha} + \hat{\beta}} \right) + \frac{\hat{\alpha}}{\hat{\alpha} + \hat{\beta}} \left( \frac{\hat{\alpha} + \hat{\beta}}{n + \hat{\alpha} + \hat{\beta}} \right).$$

## A Wrinkle: Jurisdictional Classification

- ▶ Forming groups of similar jurisdictions together can provide additional shrinkage.
- ▶ Further, forming groups in relation to model parameters will enable increased precision.
- ▶ This will yield lower errors (MSE, Variance) within the small areas.



## Jurisdictional Class Formation

- ▶ A regression was formed between direct parameter estimates and potential useful covariates.
  - ▶ Poverty, household characteristics, race/ethnicity.
- ▶ Class formations are based upon a sorted list of model predictions and their population size.
- ▶ A cumulative square-root rule on population size is used (Dalenius and Hodges, 1959).
- ▶ Jurisdictions broken into classes depending upon parameter stage and race/ethnicity model.

Finally...

**Point estimates are formed by multiplying the appropriate number of persons by the related estimated proportions and then adding over the appropriate populations.**

## Variance Computation

- ▶ Empirical Bayes methods used in point estimates.
- ▶ EB based variances tend to understate error.

Decision:

- ▶ Use Bayes methods and MCMC to estimate variances.
- ▶ EB estimates used as the location term.

$$E^{\pi(\theta|D)}[(\theta - \hat{\theta}_{\text{EB}})^2]$$

## Variance Computation

- ▶ Model priors identical to that of Gelman (2006).
- ▶ Ratio forms variances estimated via Taylor approximation.
- ▶ MCMC draws hyper-parameters which feed into model proportions.
- ▶ Those then feed into Bernoulli draws at the individual level.

## Evaluations

- ▶ Majority of evaluations performed against LEP Rate.
  - ▶ Small sample sizes for small areas are highest concern.
  - ▶ Areas that qualify via large populations will have large samples.
- ▶ Illiteracy rate estimates are above required rate for coverage.
  - ▶ Heavy model borrowing leads estimates far away from cut-off.
- ▶ Face validity checks were performed to guard against poor model performance.

## Results – ACS Weighting vs. Small Area Method

### LEP Rate Variance Reduction from Small Area Methodology by Decile Sorted By Design-Based Variance

Decile	% Reduction	
	Mean	Median
1 <sup>st</sup>	61.15	83.16
2 <sup>nd</sup>	61.45	86.93
3 <sup>rd</sup>	57.91	85.05
4 <sup>th</sup>	46.18	84.31
5 <sup>th</sup>	52.91	83.99
6 <sup>th</sup>	49.76	81.22
7 <sup>th</sup>	31.12	78.01
8 <sup>th</sup>	19.33	70.09
9 <sup>th</sup>	-31.30	65.48
10 <sup>th</sup>	-180.74	56.82

## Improvement Around the '5 Percent' Criterion

Distribution of CI Widths For all Jurisdiction/LMGs Whose LEP Rate Design-Based CI Covers 5 Percent.

	Percentile				
	5 <sup>th</sup>	25 <sup>th</sup>	50 <sup>th</sup>	75 <sup>th</sup>	95 <sup>th</sup>
Design	1.414	3.947	6.742	11.270	28.877
Model	0.414	1.796	3.030	5.776	12.431

- ▶ Model shows significant CI width shrinkage.
- ▶ Implies satisfactory performance in the small areas.

## Continuing Innovation

- ▶ Technical report available (RRS #2012-02 Joyce, et. al.).
- ▶ Bayesian implementation?
- ▶ Enhancing jurisdictional classification?
- ▶ Adaptation to 5-year old census data in 2015?
- ▶ Other?



## Contact

E-mail: [patrick.m.joyce@census.gov](mailto:patrick.m.joyce@census.gov)

Section 203 data and documentation available at the Census Redistricting Office webpage.

Work performed in collaboration with staff in the Decennial Statistical Studies Division.

