REDISTRICTING AND DIFFERENTIAL PRIVACY

Michael McDonald
University of Florida
www.electproject.org ● @ElectProject
Redistricting Experience

- Since late 1980s consulted to redistricting authorities or was an expert witness in court cases in 15 states, most recently:
  - Successful challenge of Virginia’s congressional districts as racial gerrymander
  - Unsuccessful challenge of Maryland’s congressional districts as partisan gerrymander
- Co-PI of award-winning Public Mapping Project to promote public engagement and transparency in redistricting
  - Produced DistrictBuilder online mapping tool
- Currently leading team to produce accurate precinct boundaries tiling the entire country
- Authored many scholarly redistricting publications
TWO LEGAL QUESTIONS

1) Equal Population
2) Voting Rights Compliance

Investigate these questions through a Georgia case study
Equal Population

- Congressional districts must generally have *de minimus* equal total population deviations of 1 or 9 persons.
  - If a compelling state interest exists, a 1% population deviation is allowed (*Tennant v Jefferson County Commission* 2012)

- State legislative districts may have a wider 10% deviation

- Other state and federal laws may apply
Reimagining Equal Population

If States and Courts understand the affect of differential privacy on total population counts, then the strict equal population requirement for congressional districts likely relaxes.

Will synthetic data point estimates satisfy equal population standards?

Are confidence intervals needed?

**Recommendation:** Once epsilon has been chosen, publish official confidence intervals (or approximations) of population counts.
Voting Rights
The Voting Rights Act § 2

The *Gingles* Test Three Prongs

*Thornberg v Gingles* 478 U.S. 30 (1986)

1. Can a reasonably compact district with 50% or greater minority voting-age population (VAP) be drawn?
2. Is there the presence of racially polarized voting, where sufficient numbers of Whites vote against the minority community’s preferred candidate of choice to deny the minority community an opportunity to elect their candidate?
3. The Totality of the Circumstances
Gingles Test First Prong:
Drawing a 50%+ Minority VAP District
## Georgia Statewide Counts

<table>
<thead>
<tr>
<th>Population</th>
<th>Synthetic</th>
<th>Truth</th>
<th>Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Population</td>
<td>9,687,653</td>
<td>9,687,653</td>
<td>0</td>
</tr>
<tr>
<td>Non-Hispanic White VAP</td>
<td>4,242,496</td>
<td>4,242,514</td>
<td>-18</td>
</tr>
<tr>
<td>Black VAP</td>
<td>2,141,665</td>
<td>2,140,789</td>
<td>+876</td>
</tr>
<tr>
<td>Hispanic VAP</td>
<td>538,732</td>
<td>539,002</td>
<td>-270</td>
</tr>
</tbody>
</table>

Statewide synthetic total population is constrained to equal the truth, but statewide sub-population totals are not similarly constrained.
2010 Black Voting-Age Population (Single Race + Any Combination) in Georgia 2010 Precincts

2010 BVAP Population Deviation (Synthetic - Truth)
N = 2,850 Precincts
2010 Black Voting-Age Population (Single Race + Any Combination) in Georgia 2010 Precincts

N = 2,850 Precincts
2010 Black Voting-Age Population (Single Race + Any Combination) in Georgia 2010 Precincts

2010 BVAP Population Deviation (Synthetic - Truth)

2010 BVAP Truth
N = 2,850 Precincts
Shifting Blacks from Homogenous Black Communities to Homogeneous White Communities

Synthetic Data Are (Loosely) Bounded

Black VAP, Synthetic minus Truth

Precincts with Black VAP ≥ 1,000: -9,526
Precincts with Black VAP < 1,000: +10,402
(Recall, 876 Black VAP added statewide)

Could affect Gingles test first prong in that it may be more difficult to draw a 50%+ Black VAP district.
Recommendations

Recommendation #1: Spend less of the privacy budget on total population and VAP by race and ethnicity cells

Recommendation #2: Publish official confidence intervals of counts and race and ethnicity proportions for graduated population sizes.

Recommendation #3: Apply negative spatial correlation to differential privacy algorithm to mitigate random chance of positively correlated groupings of population deviations.
Gingles Test Second Prong: Racially Polarized Voting
Secret ballot laws protect the confidentiality of individual voters

Aggregate election results reported within precincts

When precinct boundaries are known, precinct boundaries can be spatially joined to census geography, so that census population counts can be aggregated within precincts

Estimate individual voting propensities by race from these aggregate data. In the crudest form, a simple correlation
Typical Ecological Inference Analysis - 2010 Georgia Precincts Statewide
Simple Goodman’s Regression

TwoParty Vote_{Minority Preferred Candidate} = \beta_1 \text{BVAP} + \beta_2 (1 - \text{BVAP})

\beta_1 = \text{Black Vote for Minority Preferred Candidate}
\beta_2 = \text{Non-Black Vote for Minority Preferred Candidate}

Two-Stage Goodman’s Regression (controls for differential turnout rates)

Turnout as Percentage of VAP = \beta_1 \text{BVAP} + \beta_2 (1 - \text{BVAP})
Candidate's Vote as Percentage of VAP = \beta_3 \text{BVAP} + \beta_4 (1 - \text{BVAP})

\frac{\beta_3}{\beta_1} = \text{Black Vote for Minority Preferred Candidate}
\frac{\beta_4}{\beta_2} = \text{Non-Black Vote for Minority Preferred Candidate}

RxС Bayesian method proposed by Gary King and co-authors
### 2010 Governor Statewide

#### Goodman's Regression

<table>
<thead>
<tr>
<th></th>
<th>Support for Candidate of Choice</th>
<th>Standard Error</th>
<th>Diff. (Synthetic - Truth)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Truth</td>
<td>Synthetic</td>
<td>Truth</td>
</tr>
<tr>
<td>Black</td>
<td>1.0419</td>
<td>1.0443</td>
<td>0.0023</td>
</tr>
<tr>
<td>Non-Black</td>
<td>0.2017</td>
<td>0.1981</td>
<td>-0.0036</td>
</tr>
</tbody>
</table>

#### Two-Stage Goodman's Regression

<table>
<thead>
<tr>
<th></th>
<th>Support for Candidate of Choice</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Truth</td>
</tr>
<tr>
<td>Black</td>
<td>1.0861</td>
</tr>
<tr>
<td>Non-Black</td>
<td>0.1872</td>
</tr>
</tbody>
</table>

#### RxC EI

<table>
<thead>
<tr>
<th></th>
<th>Support for Candidate of Choice</th>
<th>Standard Error</th>
<th>Diff. (Synthetic - Truth)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Truth</td>
<td>Synthetic</td>
<td>Truth</td>
</tr>
<tr>
<td>Black</td>
<td>0.9510</td>
<td>0.9414</td>
<td>-0.0096</td>
</tr>
<tr>
<td>Non-Hispanic White</td>
<td>0.1929</td>
<td>0.1967</td>
<td>0.0038</td>
</tr>
<tr>
<td>Other</td>
<td>0.5884</td>
<td>0.5616</td>
<td>-0.0268</td>
</tr>
</tbody>
</table>
### 2010 Congressional District 12

<table>
<thead>
<tr>
<th></th>
<th>Support for Candidate of Choice</th>
<th>Standard Error</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Difference</td>
<td>Difference</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Synthetic (Synthetic - Truth)</td>
<td>Synthetic (Synthetic - Truth)</td>
<td></td>
</tr>
<tr>
<td><strong>Goodman's Regression</strong></td>
<td>[Table Data]</td>
<td>[Table Data]</td>
<td></td>
</tr>
<tr>
<td>Black</td>
<td>1.0302</td>
<td>0.0201</td>
<td>0.0143</td>
</tr>
<tr>
<td>Non-Black</td>
<td>0.2411</td>
<td>-0.0161</td>
<td>0.0102</td>
</tr>
</tbody>
</table>

| **Two-Stage Goodman's Regression** | [Table Data] | [Table Data] | [Table Data] |
| Black              | 1.0516                          | 0.0123         |        |
| Non-Black          | 0.2167                          | -0.0123        |        |

| **RxC EI** | [Table Data] |
| Black      | 0.9589                          | -0.0086        | 0.0205 | 0.0232 | 0.0026 |
| Non-Hispanic White | 0.2440                          | -0.0194        | 0.0172 | 0.0208 | 0.0036 |
| Other      | 0.6156                          | 0.0601         | 0.3046 | 0.2325 | -0.0721 |
Estimates are sensitive to the application of the differential privacy algorithm in the two cases examined.

More complex model specifications tend to be more sensitive to addition of measurement error.

**Recommendation #1:** Run multiple model specifications to diagnose potential problems

**Recommendation #2:** If expected simulation variance is known (or an estimate provided), apply multiple imputation methods to check sensitivity of results
REDISTRICTING AND DIFFERENTIAL PRIVACY

Michael McDonald
University of Florida
www.electproject.org • @ElectProject