

# EI Considerations

Gabe Schoenbach and Amy Becker  
MGGG Redistricting Lab | Nov. 2020

# EI Overview

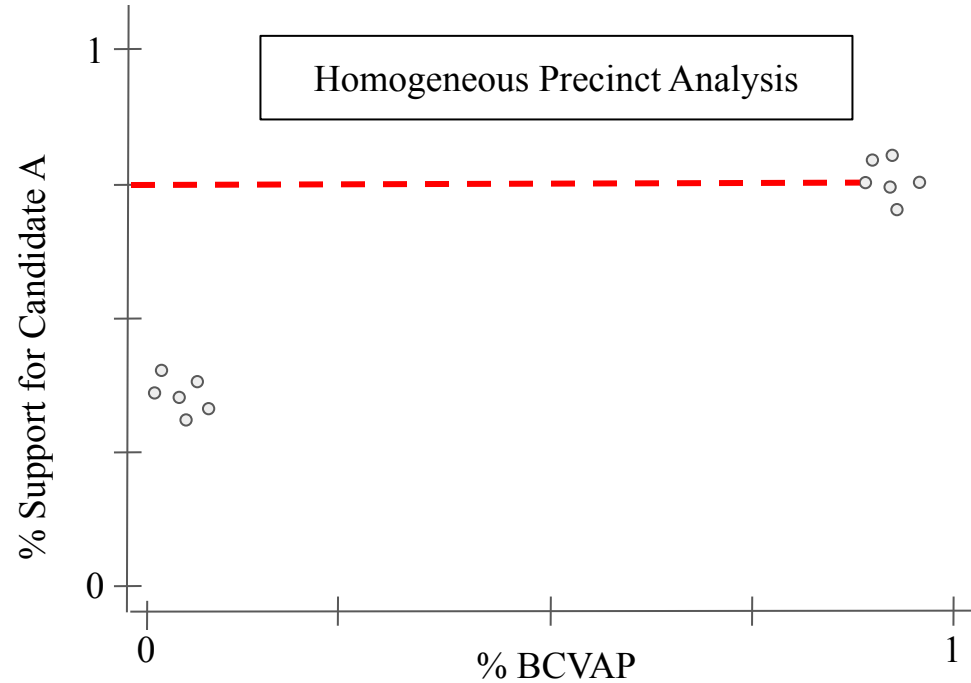
What % of Black voters chose Candidate A?  
(e.g. for VRA)

For each precinct we have:

	Candidate A	Candidate B	Totals
<b>BCVAP</b>	?	?	2000
<b>non-BCVAP</b>	?	?	3000
<b>Totals</b>	1000	800	

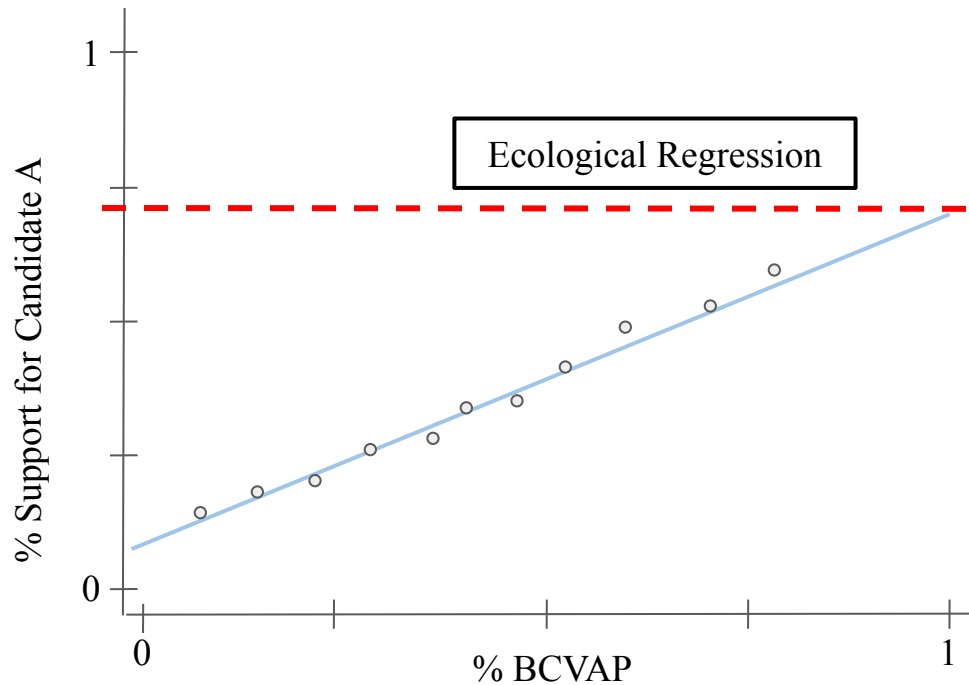
# EI Overview

(Look for patterns across all the precincts)



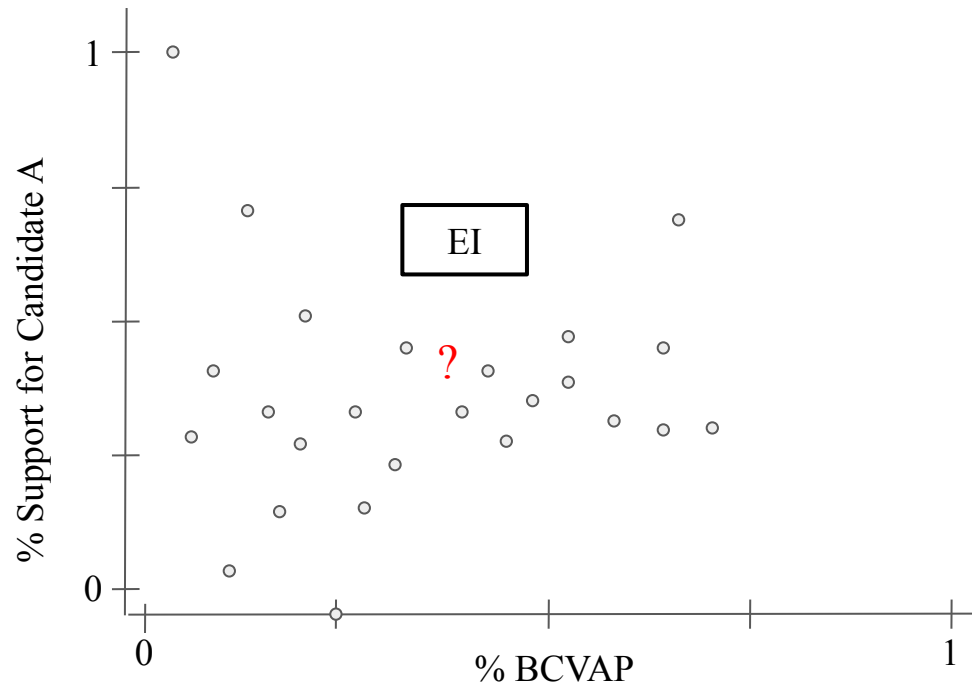
# EI Overview

(Look for patterns across all the precincts)



# EI Overview

(Look for patterns across all the precincts)



# EI Overview

- “EI” is not a single well-defined technique! Lots of flavors:
  - King’s EI, Rosen, Greiner-Quinn, ...
  - Underlying distributions (truncated bivariate normal, univariate gamma, multinomial dirichlet)
  - How to address turnout issues (e.g. Double equation vs Single equation, ‘abstain’ columns, etc)
  - Different implementations (e.g. eiPack, pyEI)
  - Lots of parameter knobs to adjust (tuning, mcmc specs, etc)
  - 2x2 vs RxC tables
  - Counts vs Proportions (each have issues); CVAP/VAP/voter file/exit polls
- EI techniques are not always reliable
  - depend on underlying assumptions that do not hold well in many scenarios
- Most practitioners likely treat EI as a “black box” without needing to understand inner-workings
- Reasonable and innocuous-seeming choices can have significant impacts!

# Our EI Method

<i>precinct <math>p_i</math></i>	Candidate 1	Candidate 2	Neither
BCVAP			
HCVAP			
WCVAP			
OCVAP			

Use 4 demographic groups

# Our EI Method

<i>precinct <math>p_i</math></i>	Candidate 1	Candidate 2	Neither
BCVAP			
HCVAP			
WCVAP			
OCVAP			

Use 4 demographic groups

Aggregate precincts statewide



# Our EI Method

<i>precinct <math>p_i</math></i>	Candidate 1	Candidate 2	Neither
BCVAP			
HCVAP			
WCVAP			
OCVAP			

Use 4 demographic groups

Aggregate precincts statewide

Immediately infer turnout

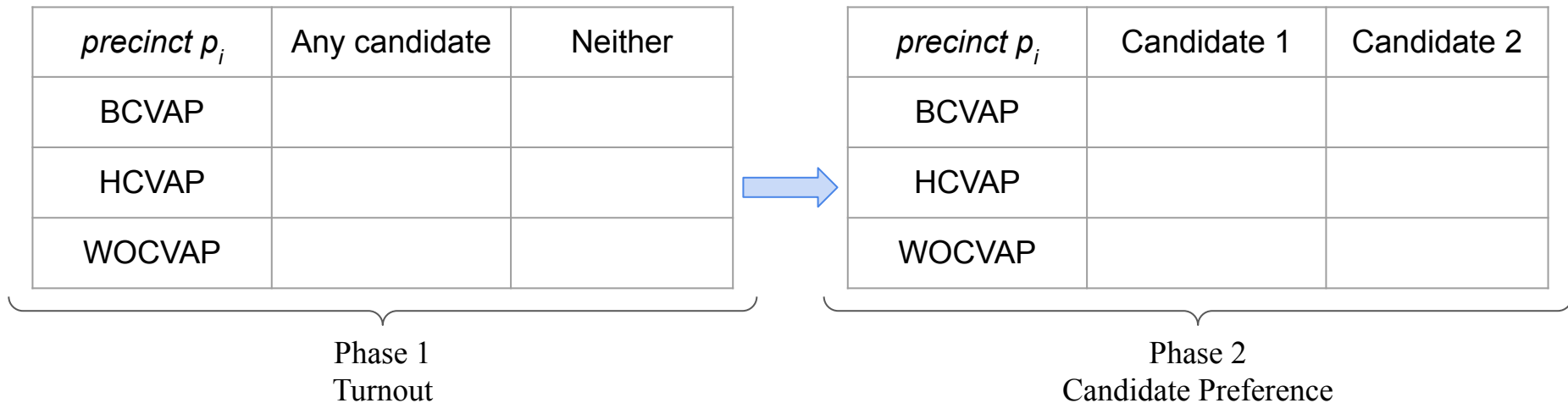
# Our EI Method

<i>precinct <math>p_i</math></i>	Candidate 1	Candidate 2	Neither
BCVAP			
HCVAP			
WCVAP			
OCVAP			

## Toggles

Use 4 demographic groups	Aggregate precincts statewide	Immediately infer turnout
Combine White + Other	Aggregate precincts countywide	Make a turnout phase

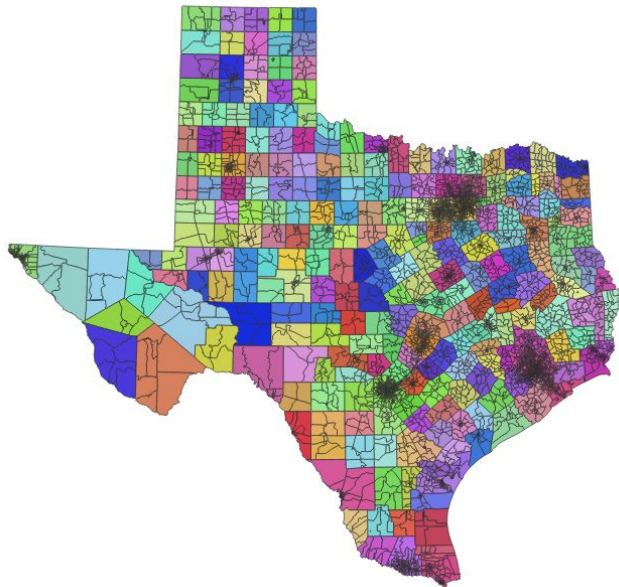
# Chen's EI Method



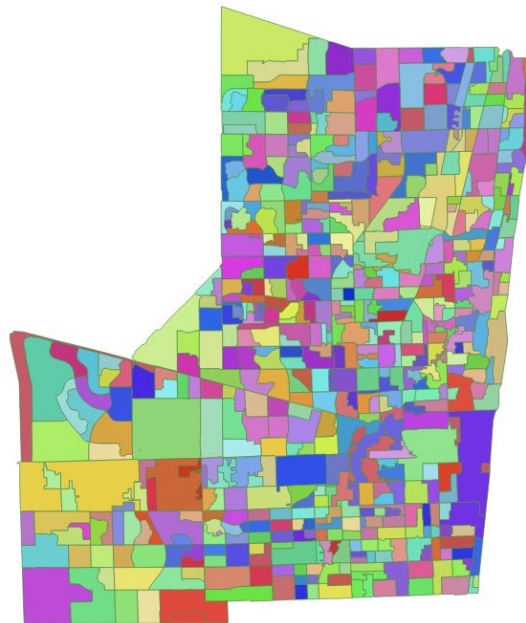
## Toggles

Use 4 demographic groups	Aggregate precincts statewide	Immediately infer turnout
Combine White + Other	Aggregate precincts countywide	Make a turnout phase

# Data



Texas  
9083 precincts  
~8-14 hours/election



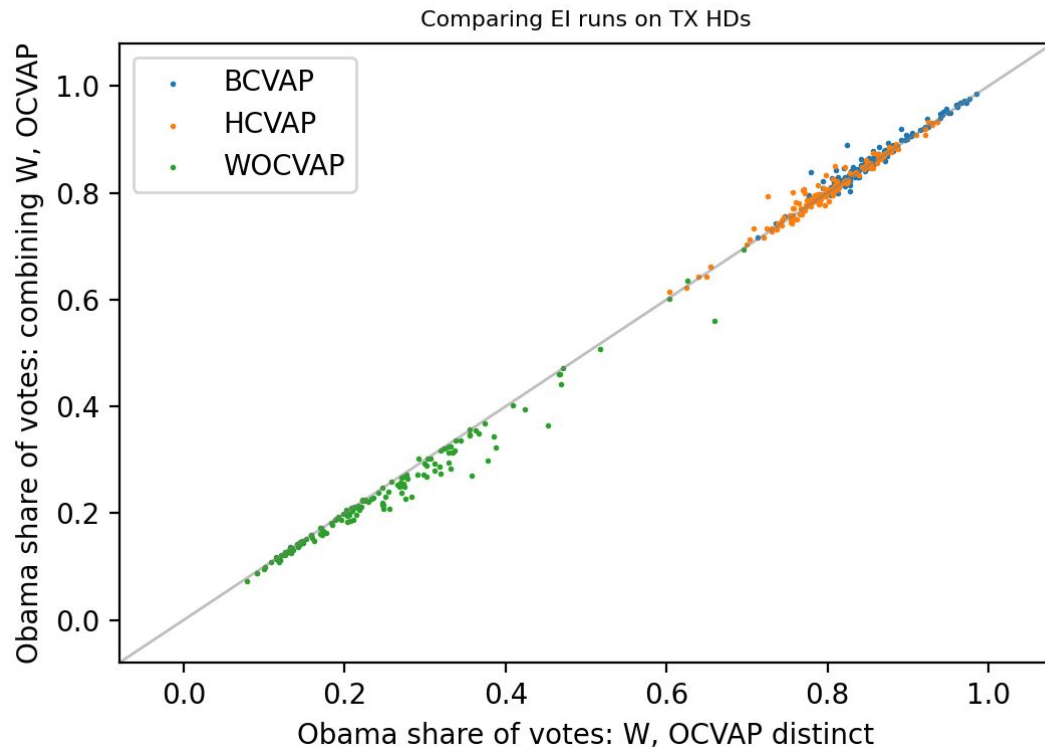
Broward Co., FL  
576 precincts  
~15 min/election

# Results: Consolidating Demographic Groups

- Election: PRES12 General
- Aggregate precinct results to Texas state house districts
- Consolidating tends to slightly underestimate WCVAP support for Obama, *relative to not consolidating*

Use 4 demographic groups

Combine White + Other

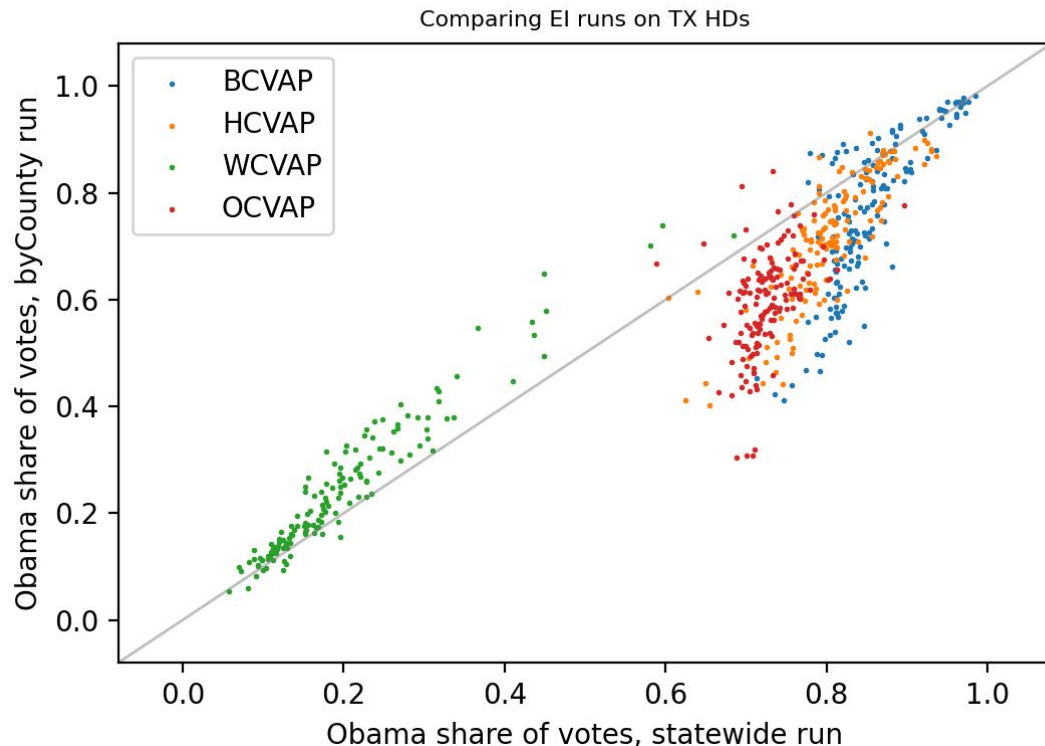


# Results: Statewide vs. County-by-County

- Election: PRES12 General
- Aggregate precinct results to Texas state house districts
- byCounty tends to overestimate WCVAP support for Obama, *relative to statewide run*
- Underestimates for BCVAP, HCVAP, OCVAP

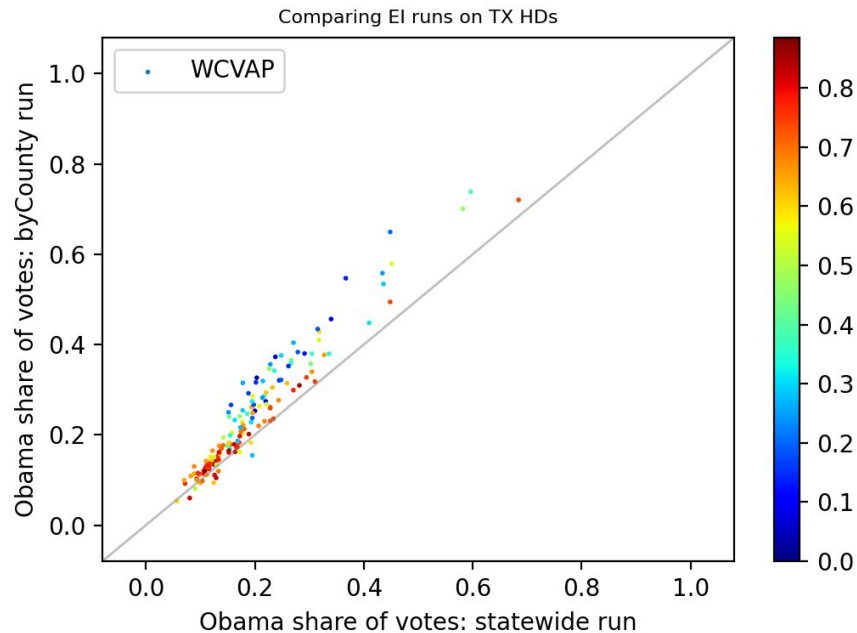
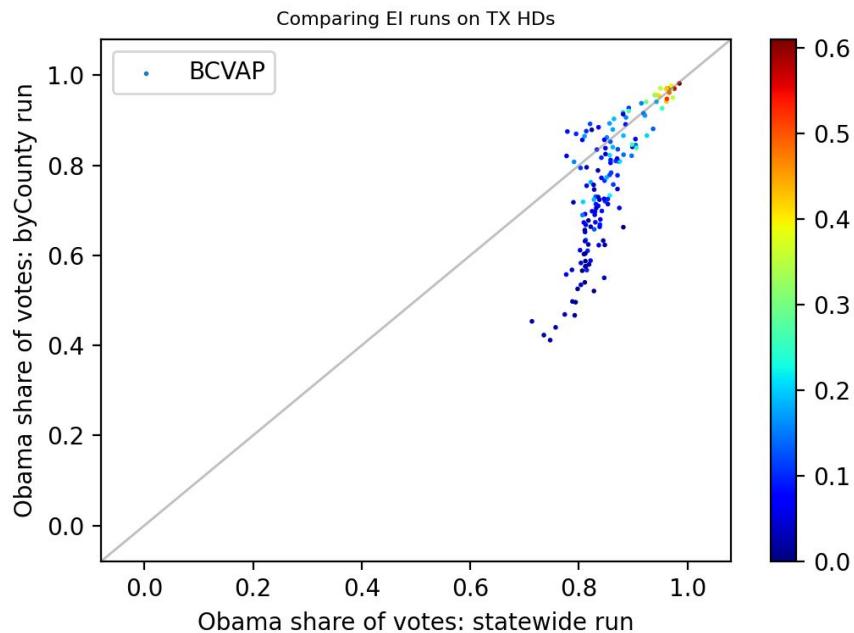
Aggregate precincts statewide

Run EI county-by-county



# Results: Statewide vs. County-by-County

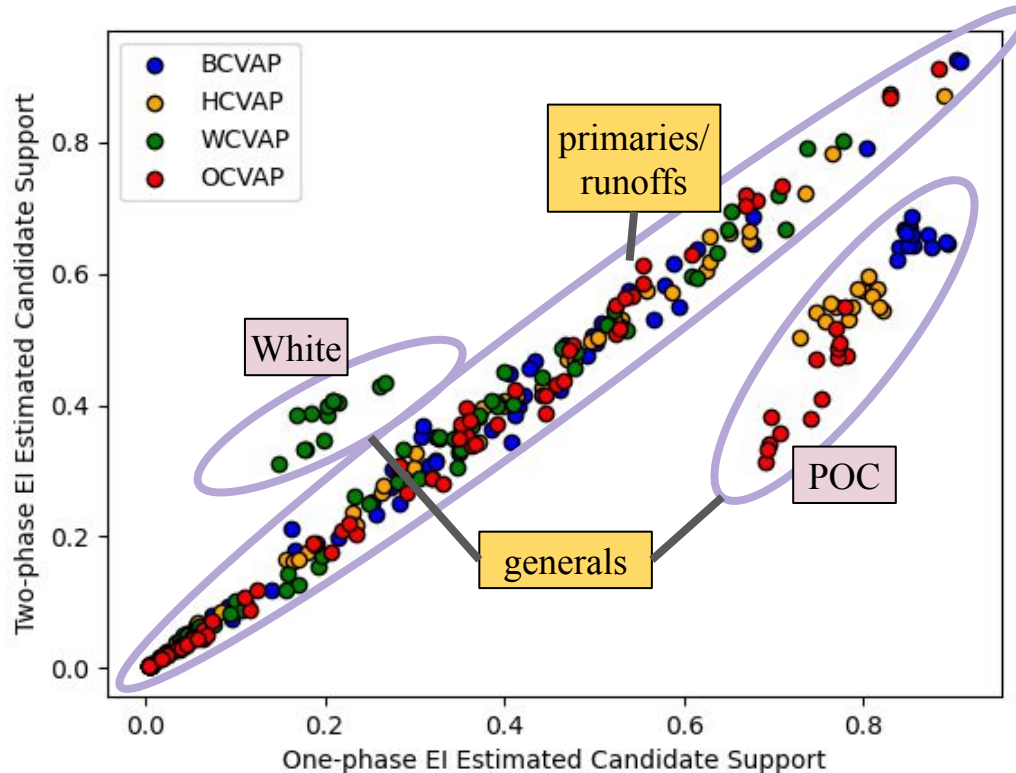
- Which HDs are most affected by this toggle?



# 1 Phase vs 2 Phase: Texas Results

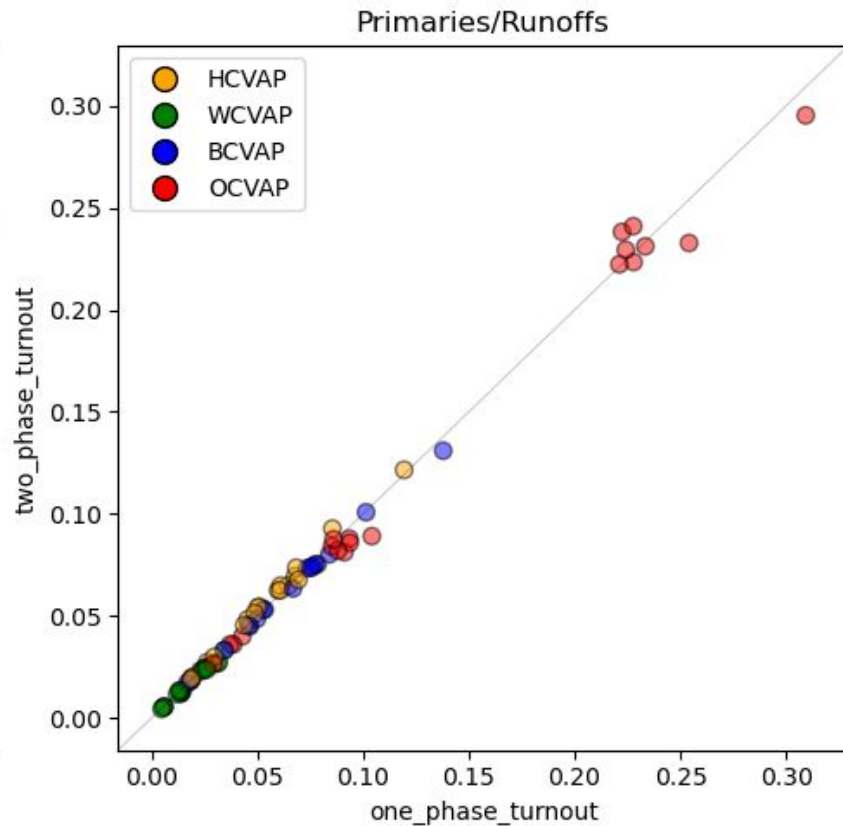
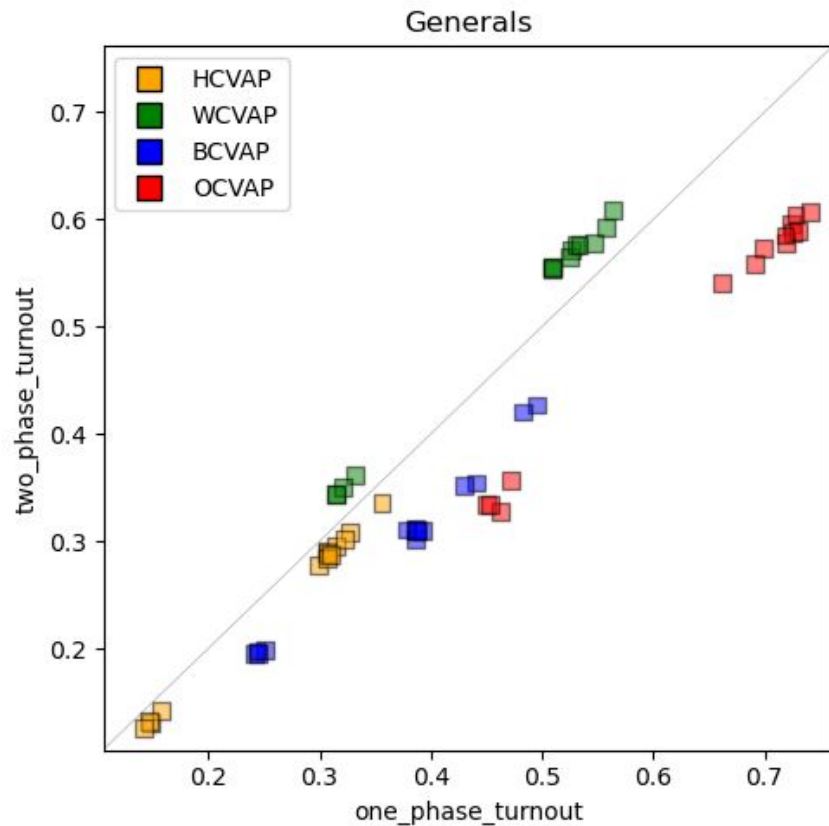
**One Phase:** Estimate candidate support and turnout in single run, using 'abstain'/'neither' column (i.e.  $R_x(C+1)$ )

**Two Phase:** First estimate turnout-by-race ( $R_x2$ ), then feed output into second run to estimate candidate support ( $R_xC$ )





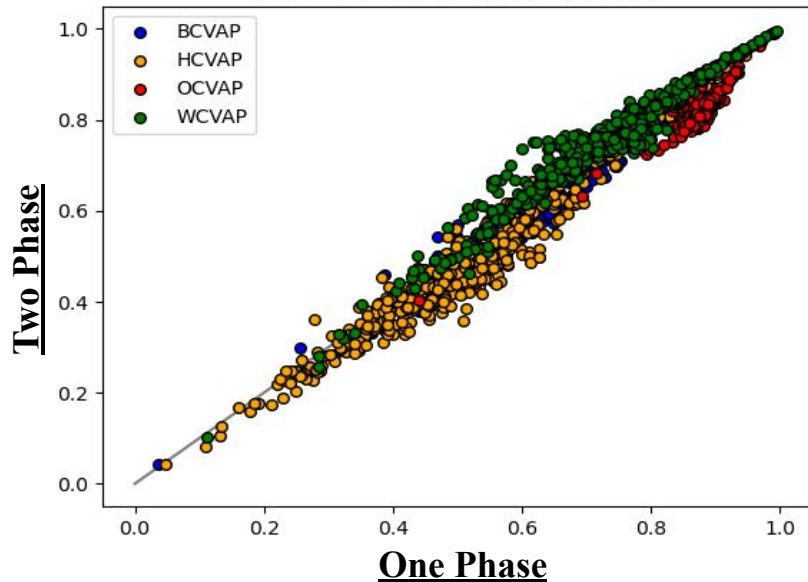
# 1 Phase vs 2 Phase: Texas Results



# 1 Phase vs 2 Phase: Broward 2016 Presidential Results

## Turnout

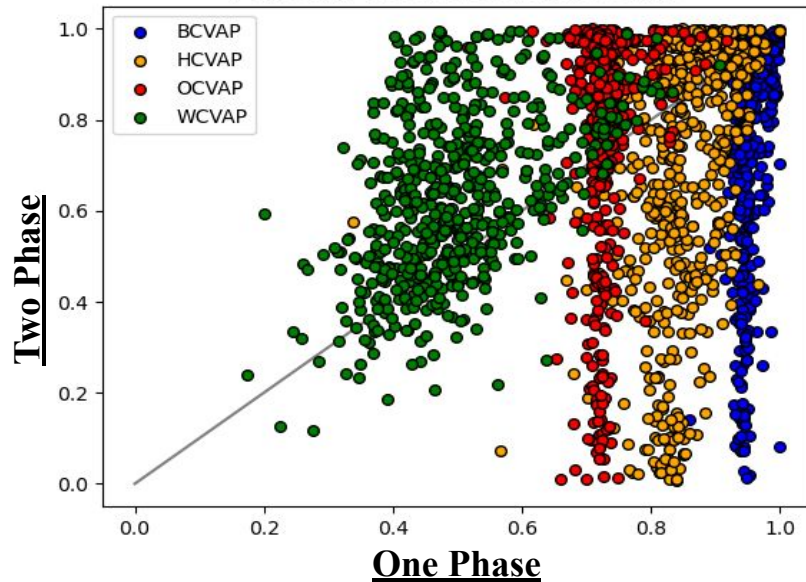
X avgs: 0.577, 0.515, 0.871, 0.726; Y avgs: 0.565, 0.499, 0.813, 0.748;  
Pearson r: 0.974, 0.973, 0.9, 0.969



	BCVAP	HCVAP	OCVAP	WCVAP
One Phase	57%	52%	87%	73%
Two Phase	57%	50%	81%	75%

## Support Clinton

X avgs: 0.968, 0.841, 0.722, 0.479; Y avgs: 0.831, 0.678, 0.793, 0.583;  
Pearson r: 0.306, 0.353, 0.146, 0.449



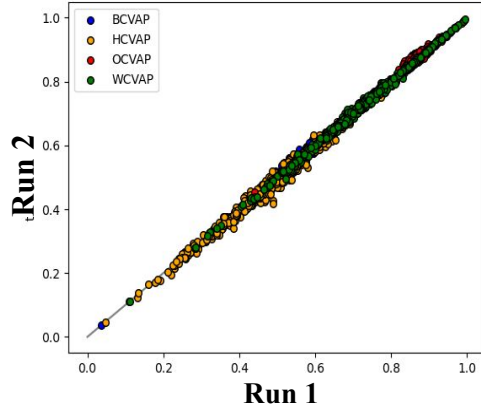
	BCVAP	HCVAP	OCVAP	WCVAP
One Phase	97%	84%	72%	48%
Two Phase	83%	68%	79%	58%

# 1 Phase vs 2 Phase: Replicability

## One Phase

### Turnout

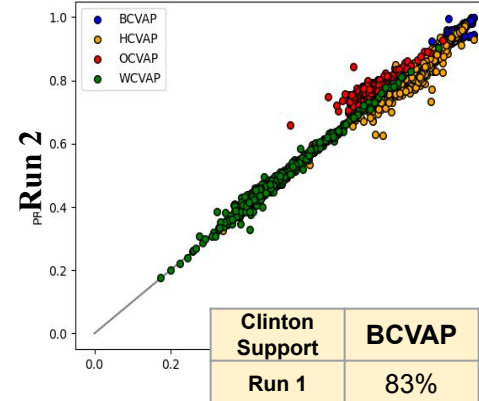
X avgs: 0.577, 0.515, 0.871, 0.726; Y avgs: 0.58, 0.51, 0.879, 0.727;  
Pearson r: 0.995, 0.997, 0.964, 0.998



	Pearson r
BCVAP	.995
HCVAP	.997
OCVAP	.964
WCVAP	.998

### Support Clinton

X avgs: 0.968, 0.841, 0.722, 0.479; Y avgs: 0.967, 0.82, 0.767, 0.482;  
Pearson r: 0.894, 0.946, 0.876, 0.994

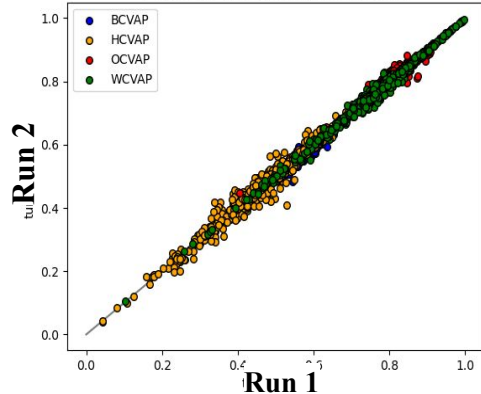


	Pearson r
BCVAP	.894
HCVAP	.946
OCVAP	.876
WCVAP	.994

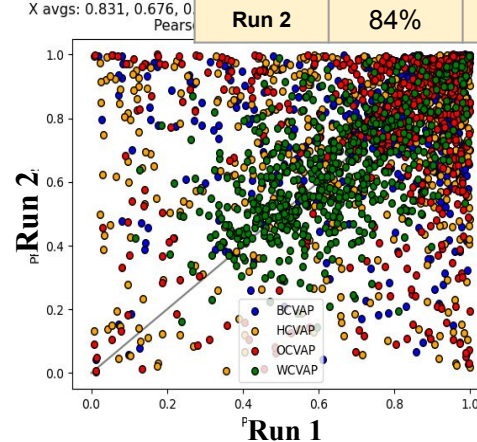
Clinton Support	BCVAP	HCVAP	OCVAP	WCVAP
Run 1	83%	68%	79%	58%
Run 2	84%	69%	78%	58%

## Two Phase

X avgs: 0.565, 0.499, 0.813, 0.748; Y avgs: 0.562, 0.502, 0.819, 0.747;  
Pearson r: 0.993, 0.992, 0.918, 0.996



	Pearson r
BCVAP	.993
HCVAP	.992
OCVAP	.918
WCVAP	.996



	Pearson r
BCVAP	.270
HCVAP	.268
OCVAP	.255
WCVAP	.555

# Next Steps

- Understand replicability issue with Two Phase
- Use FL voter file to see how ground-truth turnout-by-race compares to EI estimates
- Repeat analysis for other elections (including primaries)
- Try replicating using PyEI