A NOVEL METHOD FOR SHOWING RACIALLY POLARIZED VOTING: BAYESIAN IMPROVED SURNAME GEOCODING

MATTHEW BARRETO, MICHAEL COHEN, LOREN COLLINGWOOD, CHAD W. DUNN, & SONNI WAKNIN[∞]

UCLA VOTING RIGHTS PROJECT

ABSTRACT

Section 2 of the Voting Rights Act is one of the most important tools for litigants challenging discriminatory voting procedures. The Supreme Court outlined the test governing vote dilution claims—which are claims that an electoral system, process, or procedure weakens a minority group's ability to elect candidates of their choice—under Section 2 of the Voting Rights Act in Thornburg v. Gingles*. The Court held, in part, that Section 2 requires plaintiffs to prove that voting patterns within the challenged jurisdiction are polarized by race, which is called racially polarized voting. Put simply, racially polarized voting means that voters of different races vote cohesively with voters of their race and have opposite electoral preferences of other races. Because vote choice is private and most states do not track the race of voters, social scientists have developed statistical methods to make the evidentiary showing that* Gingles *requires. These methods are decades old and are often the subject of intense scrutiny in vote dilution trials. In some cases, the size of the jurisdiction and the quality of the voter file and voting records prevent plaintiffs from meeting their burden of proof. Analyzing the presence of racially polarized voting will be one of the most important issues during and after the redistricting cycle currently underway following the 2020 Census. Within the last year, an innovative method adapted from other fields of study called Bayesian Improved Surname Geocoding (BISG) has been applied to racially polarized voting analysis in vote dilution cases and has been approved by a federal district court in New York and the U.S. Court of Appeals for the Second Circuit. BISG has*

∞ Author names presented in alphabetical order. Authorship is equal. Initial research was supported with funding from the New York Civil Liberties Union (NYCLU). Matthew Barreto is a Professor of Political Science and Chicana/o & Central American Studies at University of California, Los Angeles (UCLA) and co-founder and Director of the UCLA Voting Rights Project (VRP). Michael Cohen is an Associate at Renne Public Law Group and a former Legal Fellow at the UCLA VRP. Loren Collingwood is an Associate Professor in the Department of Political Science at University of New Mexico. Chad W. Dunn is an instructor at the UCLA Luskin School of Public Affairs and the UCLA School of Law and is co-founder and Legal Director of the UCLA VRP. Sonni Waknin is Program Manager and Voting Rights Counsel at the UCLA VRP. The authors would like to thank Michael Pernick, Redistricting Counsel at NAACP Legal Defense & Education Fund, for his feedback, and the editors of the *N.Y.U. Review of Law & Social Change* for their editorial assistance.

received little scholarly attention in voting rights legal scholarship, but it promises to be a critical advancement in detecting vote dilution. This Article seeks to showcase this method, equipping voting rights advocates, social scientists and governments alike with additional tools to secure equal voting rights nationwide. This Article argues that BISG can be used by voting rights advocates to bolster racially polarized voting analysis when the necessary data is available and of sufficient quality. Further, BISG can be helpful to smaller jurisdictions which might have smaller sample sizes in American Community Survey data or a smaller number of precincts.

I.
INTRODUCTION

The 2021–22 redistricting of the United States could be one of the more racially discriminatory redistricting cycles since the passage of the federal Voting Rights Act (VRA) in 1965.[1] Voting rights advocates face this risk amid significant and persistent racial differences in voting patterns.[2] While courts have continued to use decades-old social science methods to support legal challenges to vote

1. *See* MICHAEL C. LI, BRENNAN CTR. FOR JUST., THE REDISTRICTING LANDSCAPE, 2021–22, 3 (2021), https://www.brennancenter.org/sites/default/files/2021-02/2021_2_11_State%20of%20Redistricting.pdf [https://perma.cc/H74S-QZ43] (explaining how the Supreme Court has gutted the core protections of the federal Voting Rights Act and permitted political gerrymanders).

2. *See* Kevin Morris, *Large Racial Turnout Gap Persisted in 2020 Election,* BRENNAN CTR. FOR JUST. (Aug. 6, 2021), https://www.brennancenter.org/our-work/analysis-opinion/large-racial-turnout-gap-persisted-2020-election [https://perma.cc/MJ8F-F4K3]. *See also* Bernard L. Fraga, *The Turnout Gap Between Whites and Racial Minorities is Larger Than You Think—and Hard to Change,* WASH. POST. (Sept. 25, 2018), https://www.washingtonpost.com/news/monkey-cage/wp/2018/09/25/the-turnout-gap-between-whites-and-racial-minorities-is-larger-than-you-think-and-hard-to-change/ [https://perma.cc/2S76-SWGH].

dilution,[3] practitioners on all sides can benefit from updated and advanced methods. Within the last year, one trusted method to support legal challenges to vote dilution has received judicial approval: Bayesian Improved Surname Geocoding (BISG).[4] Although BISG has received significant scholarly attention over the last decade as a highly reliable way to estimate race across a number of disciplines and applications, BISG's more recent application to estimating the race of voters, aggregated down to precinct counts, promises to be a critical advancement in evaluating vote dilution. This Article seeks to explain the BISG methodology with some practical examples and offer guidance to voting rights advocates, social scientists, and state and local governments in their effort to secure equal voting rights nationwide.

Today, racial, ethnic, and language minority[5] communities lack critical protections that previously shielded them from discriminatory voting laws, practices, and procedures.[6] As enacted, the VRA included a preclearance requirement in its Section 5, which prohibited jurisdictions with a history of

3. *See, e.g.*, Bernard Grofman & Samuel Merrill, *Ecological Regression and Ecological Inference*, *in* ECOLOGICAL INFERENCE: NEW METHODOLOGICAL STRATEGIES 123–24 (Gary King, Ori Rosen, & Martin A. Tanner eds., 2004) (stating that the methodologies that serve as the foundation for ecological inference were founded in the 1950s, which were then advanced into the ecological inference method now used today by Gary King in 1997); Perez v. Abbott, No. SA-11-CV-360, 2017 mulat *34 (W.D. Tex. Apr. 20, 2017) (discussing how one of the parties' experts used multivariate analysis to determine if there was racial polarization in voting, but noting that the differences between multivariate and bivariate analysis are often exaggerated), *rev'd*, 138 S.Ct. 2305 (2018); Matt Barreto, Loren Collingwood, Sergio Garcia-Rios, & Kassra AR Oskooii, *Estimating Candidate Support in Voting Rights Act Cases: Comparing Iterative EI and EI-R×C Methods*, 51 SOC. METHODS & RSCH. 271, 273 (2019) (comparing the efficacy of using two dominant forms of ecological inference, a method which helps assess whether racial bloc voting has occurred during an election in voting rights cases).

4. *See* NAACP Spring Valley Branch v. E. Ramapo Cent. Sch. Dist., 462 F. Supp. 3d 368, 383–84 (S.D.N.Y. 2020); Clerveaux v. E. Ramapo Cent. Sch. Dist., 984 F.3d 213, 233–37 (2d Cir. 2021); *see also infra* Part IV.C.

5. The term "language minority" or "language minority group" is defined by 52 U.S.C. § 10310(c) to mean persons who are "American Indian, Asian American, Alaskan Natives or of Spanish heritage." Sometimes this Article uses the term "minority" to represent historically marginalized racial, ethnic, and language communities. Although the term "minority" may be imprecise and "viewed pejoratively," especially in reference to people of color, *see, e.g.*, *Racial and Ethnic Identity*, APA STYLE (Sept. 2019), https://apastyle.apa.org/style-grammar-guidelines/bias-free-language/racial-ethnic-minorities [https://perma.cc/KW4M-SVN3], "minority" is the term used in vote dilution cases and is consistent with statutory language, *see, e.g.*, Thornburg v. Gingles, 478 U.S. 30 (1986).

6. *See* Shelby County v. Holder, 570 U.S. 529 (2013); Brnovich v. Democratic Nat'l Comm., 141 S. Ct. 2321 (2021). *See also* Richard L. Hasen, *The Supreme Court's Latest Voting Rights Opinion Is Even Worse than It Seems,* SLATE (July 8, 2021), https://slate.com/news-and-politics/2021/07/supreme-court-sam-alito-brnovich-angry.html [https://perma.cc/V97T-P8LV] ("[As a result of *Brnovich*,] states can put up roadblocks to minority voting and engage in voter suppression with few legal consequences once a state has raised tenuous and unsupported concerns about the risk of voter fraud. It's exactly the opposite of what Congress intended when it strengthened Section 2 of the Voting Rights Act in 1982, and it turns on its head the 'non-retrogression' principle that Congress wrote in Section 5 of the act and that the court essentially killed off eight years ago in *Shelby County v. Holder*.").

discrimination from implementing any change affecting voting without receiving preapproval from the U.S. Attorney General or the U.S. District Court for the District of Columbia.[7] The provision was meant to thwart efforts to disenfranchise voters by requiring the federal government to confirm that a proposed change in voting practices would not discriminate against protected minorities in violation of the VRA *before* the changes took effect.[8] In 2013, *Shelby County v. Holder* decimated the preclearance regime by striking down the triggering formula in Section 4(b) that determined which areas of the country were required to undergo preclearance.[9] In short, by invalidating Section 4(b)'s formula, the Supreme Court made Section 5's preclearance requirement inoperable (unless and until Congress enacts a new formula that withstands judicial scrutiny). Preclearance was critically important for protecting voting rights in communities that lived in covered jurisdictions for two reasons. First, under the preclearance regime, state or local governments had the evidentiary burden of proving to the federal government that a voting practice did not harm voters of color in the jurisdiction that was enacting the new voting restriction before the change took effect.[10] Second, Section 5 set a standard that new districting plans could not weaken, move back, or otherwise "retrogress" the political representation of communities in covered jurisdictions.[11] After *Shelby County*, plaintiffs now must affirmatively bring Section 2 challenges in court, and they bear the evidentiary burden in all federal court cases seeking a remedy for vote dilution under Section 2 of the VRA.[12] The redistricting that occurred during 2021−2022 is the first round of redistricting to lack Section 5's protections since 1971.[13]

With the end of the preclearance regime for covered jurisdictions, plaintiffs choosing to challenge a redistricting scheme or method of election for vote dilution

7. 52 U.S.C. § 10304, *invalidated by* Shelby County v. Holder, 570 U.S. 529 (2013).

8. *Id. See also* Robert A. Kengle & Marcia Johnson-Blanco, *What Is Next for Section 5 of the Voting Rights Act?*, 39 HUM. RTS. 6, 6 (2012) ("Section 5 frequently is called one of the most effective civil rights laws ever passed by Congress, and it has been crucial to the historic political empowerment of minority voters in the South and Southwest. Since it was enacted in 1965, section 5 has led to thousands of racially discriminatory voting changes being blocked before they could be put into practice.").

9. 570 U.S. at 557.

10. *See* 52 U.S.C. §§ 10303(a)(1), 10304.

11. Retrogression of a minority performing district means that a redistricting plan or voting practice "would lead to a retrogression in the position of racial minorities with respect to their effective exercise of the electoral franchise." Georgia v. Ashcroft, 539 U.S. 461, 466 (2003) (quoting Beer v. United States, 425 U.S. 130, 141 (1976)).

12. Section 2 of the Voting Rights Act does not have a burden shifting mechanism and requires plaintiffs to prove a violation by the "totality of circumstances." *See* 52 U.S.C. § 10301.

13. Congress has not passed a new coverage formula, meaning that Section 5 is still inoperable after *Shelby County,* 570 U.S. 529 (2013). Thus, states and jurisdictions that were required to submit to preclearance under the law are not required to do so during 2021 redistricting. *See* LI, *supra* note 1, at 9–10.

under the VRA only have Section 2.[14] Vote dilution refers to the discriminatory effect that districting or methods of elections have when they work "to cancel out or minimize" voting strength.[15] Vote dilution concerns the ability of persons to have their votes weigh or count equally to others, especially in the population composition of a district.[16] Drawing election districts in ways that improperly weaken a group's voting power and/or ability to elect candidates of choice can result in vote dilution.[17] Although minority vote dilution can manifest in many ways, two typical forms involve "cracking" and "packing." "Cracking" involves spreading minority voters across multiple districts to reduce their influence. [18] "Packing" involves making minority voters a super-majority in a smaller number of districts, thus reducing the group's electoral influence.[19] Vote dilution can also occur through the use of at-large election systems.[20] At-large election systems are characterized by a multi-member governing body in which each member is elected by the whole population of a jurisdiction rather than geographic districts.[21] At-large systems can result in vote dilution because even though a community's minority citizen voting age population can be sizable and vote cohesively, it may not have enough representation in the wider electorate to elect a candidate of choice in a jurisdiction-wide majority vote.[22] In at-large systems, the minority is only ever able to elect candidates of choice if the majority population or group decides to vote for minority-preferred candidates.[23] Vote dilution also occurs when a redistricting body fails to draw a majority-minority or coalition district

14. Plaintiffs may use other legal theories to challenge racial vote dilution and discriminatory redistricting schemes based on the specific facts of a particular case. For example, plaintiffs can bring claims under the Fourteenth and/or Fifteenth Amendments, including intentional race discrimination, *see* York v. City of St. Gabriel, 89 F. Supp. 3d 843, 850 (M.D. La. 2015), or racial gerrymandering, often called a *Shaw* claim, *see* Shaw v. Reno, 509 U.S. 630 (1993); Miller v. Johnson, 515 U.S. 900 (1995). Plaintiffs can also bring malapportionment challenges, *see* Reynolds v. Sims, 377 U.S. 533 (1964), or state law claims.

15. *See* 29 C.J.S. *Elections* § 91 (Year).

16. *Id.*

17. *Redistricting Definitions*, UCLA LATINO POL'Y & POL. INITIATIVE, https://latino.ucla.edu /redistricting-definitions/ [https://perma.cc/C6GC-B6B2] (last visited Oct. 14, 2021). In this Article, we address vote dilution with respect to minority voting power, but vote dilution can also occur based on other group characteristics, such as partisan preference.

18. *Id.*

19. *Id.*

20. *See* Barbara L. Berry & Thomas R. Dye, *The Discriminatory Effects of At-Large Elections,* 7 FL. STATE U. L. REV. 85 (1979); *see also*, *At-Large Election Systems*, FAIRVOTE, http://archive .fairvote.org/?page=766 [http://perma.cc/88TS-A4H8] (last updated Dec. 2009) (Because at-large districts "allow 50 percent of voters to control 100 percent of seats, [they] . . . typically result in racially and politically homogenous elected bodies."). Prior to *Shelby County*, the DOJ regularly blocked proposals to switch from single-member to at-large districts in *Texas. Patino v. City of Pasadena*, 230 F. Supp. 3d 667, 697 (S.D. Tex. 2017). *See* Edward K. Olds, *More Than "Rarely Used": A Post-Shelby Judicial Standard for Section 3 Preclearance*, 117 COLUM. L. REV. 2185, 2223 (2017).

21. *See* Berry & Dye, *supra* note 20, at 86.

22. *See id.* at 88.

23. *See id.* at 86, 88.

when it is required by Section 2 of the Voting Rights Act.[24] Among other things, to win a vote dilution claim under Section 2, the plaintiff must prove that the relevant minority group "is politically cohesive . . . [and] that the white majority votes sufficiently as a bloc to enable it . . . usually to defeat the minority's preferred candidate."[25] The evidence offered by social science experts to meet this evidentiary standard is referred to by practitioners and courts alike as "racially polarized voting" analysis.[26]

Because census data is not reported at the level of individual voters, and voter files kept by many states do not contain racial or ethnic demographics of each voter,[27] social scientists have developed statistical methods to estimate voters' racial information.[28] Under certain circumstances, these methods face limitations. Different racial groups register and vote at different rates depending on the circumstances surrounding an election and other intersectional identities of a

24. *See* League of United Latin Am. Citizens v. Perry, 548 U.S. 399 (2006) [hereinafter *LULAC*]. In *LULAC*, the Supreme Court found that Section 2 of the Voting Rights Act was violated when Texas implemented a redistricting plan that "took away the Latinos' opportunity [to elect candidates of choice] because Latinos were about to exercise it." *Id.* at 440.

25. Thornburg v. Gingles, 478 U.S. 30, 50–51 (1986).

26. *See generally* M.V. Hood III, Peter A. Morrison, & Thomas M. Bryan, *From Legal Theory to Practical Application: A How-To for Performing Vote Dilution Analyses*, 99 SOC. SCI. Q. 536, 536–52 (2018) (discussing use of ecological inference to conduct racially polarized voting analysis, including differences between Census, voter file analysis, and BISG to measure voter race); Kareen U. Crayton, *Sword, Shield, and Compass: The Uses and Misuses of Racially Polarized Voting Studies in Voting Rights Enforcement,* 64 RUTGERS L. REV. 973, 973–1018 (2012) (discussing different ways to measure racially polarized voting including exit polls, homogenous precinct analysis, and ecological inference); Christopher S. Elmendorf, Kevin M. Quinn, & Marisa A. Abrajano, *Racially Polarized Voting*, 83 U. CHI. L. REV. 587, 589 (2016) ("At the heart of vote dilution law is the concept of racially polarized voting.").

27. Kevin Deluca & John A. Curiel, Validating the Applicability of BISG to Congressional Redistricting 1, 3 (unpublished manuscript), https://electionlab.mit.edu/sites/default/files/2021-07/deluca-curiel_validating_bisg.pdf [https://perma.cc/W6J9-8RYN] ("Many states do not collect individual race data in their voter files–including states like Texas, Pennsylvania, and Wisconsin, which are often subjects of contentious gerrymandering litigation.") (last visited Oct. 14, 2021); U.S. ELECTION ASSISTANCE COMM'N, *Availability of State Voter File and Confidential Information* (2020), https://www.eac.gov/sites/default/files/voters/Available_Voter_File_Information.pdf [https://perma.cc/TY5U-WSWE] (listing the information that each voter file contains, including states that provide racial demographic information)*; Commercial Voter Files and the Study of U.S. Politics*, PEW RSCH. CTR. (Feb. 15, 2018), https://www.pewresearch.org/methods/2018/02/15/demographic-data [https://perma.cc/R5US-SZ52] ("In 16 states or portions of states, largely in the South, the Voting Rights Act of 1965 mandated that states list voters' race on the state voter rolls. However, in states where this information is not available, vendors attempt to use information from other sources such as identifying common surnames or if someone lives in an area that is densely populated by a particular race.").

28. *See* Grofman & Merrill, *supra* note 3 (describing multivariate analysis of voting patterns); Loren Collingwood, Kassra Oskooii, Sergio Garcia-Rios, & Matt Barreto, *eiCompare: Comparing Ecological Inference Estimates across EI and EI:RxC,* 8 R J. 92 (2016) (comparing two common ecological inference methods).

group, such as age and language.[29] In jurisdictions with similar rates of registration and turnout across racial groups, existing methods of showing racially polarized voting likely remain highly accurate. In places where there is a particular imbalance in turnout rates, however, conventional methods may not be precise enough to speak to the race of those who actually voted.[30] BISG works to fill these data gaps.[31] BISG relies on the voter file and vote history data, which is usually accessible to analysts and in many cases already digitized and ready for analysis.[32] In other words, it harnesses widely available information on actual voters to understand voting patterns with more precision.

Analyzing the presence of racially polarized voting will be one of the most important issues during and after the 2021−2022 redistricting round. Voting rights plaintiffs will need to demonstrate racially polarized voting to prevail in their vote dilution claims,[33] and governments must redistrict in compliance with the VRA.[34] BISG is a critical innovation in providing the precision demanded by both processes.

This Article describes the evidentiary burden required to prove racially polarized voting in vote dilution cases, some of the challenges in meeting this burden presented by conventional methods in some cases, and how the BISG method can help demonstrate meet this burden under certain circumstances. Part II of this Article provides background on Section 2 of the VRA and outlines the legal framework currently used for Section 2 cases. Part III identifies the potential limitations of current methods used to show racially polarized voting. In Part IV, this Article makes the case for utilizing BISG. It demonstrates why BISG is an important tool for redistricting and subsequent voting rights litigation, and it documents recent judicial approval of the method. The Article concludes that, while courts should not *require* BISG, they should continue to accept it, and expert witnesses and voting rights advocates should consider using BISG as an additional method to bolster racially polarized voting analysis when the necessary data is available and of sufficient quality.

29. *See* John R. Logan, Jennifer Darrah, & Sookhee Oh, *The Impact of Race and Ethnicity, Immigration and Political Context on Participation in American Electoral Politics*, 90 SOC. FORCES 993, 1010–11 (2012).

30. *See, e.g.*, Stephen P. Klein, Jerome Sacks, & David A. Freedman, *Ecological Regression Versus the Secret Ballot*, 31 JURIMETRICS 393 (1991).

31. *See infra* Part IV.

32. *See* Berry & Dye, *supra* note 20, at 3.

33. *See* Thornburg v. Gingles, 478 U.S. 30, 52–74 (1986).

34. *See* U.S. DEPT. OF JUST., GUIDANCE UNDER SECTION 2 OF THE VOTING RIGHTS ACT, 52 U.S.C. 10301, FOR REDISTRICTING AND METHODS OF ELECTING GOVERNMENT BODIES 1–2 (2021), https://www.justice.gov/opa/press-release/file/1429486/download [https://perma.cc/6DB4-D9WV].

II.
VOTING RIGHTS ACT § 2: THE LEGAL FRAMEWORK AND PRACTICAL IMPLICATIONS

Congress designed the Voting Rights Act of 1965 "to banish the blight of racial discrimination in voting, which has infected the electoral process in parts of our country for nearly a century."[35] Section 2 of the Voting Rights Act provides a cause of action[36] to challenge a state's use of voter qualifications, standards, practices, or procedures that result in the denial or abridgement of the right to vote "on account of race or color."[37] Section 2 can be used to challenge both vote denial and vote dilution.[38] Vote denial occurs when the ability of voters to cast a ballot or have that ballot counted is impeded.[39] These cases concern the ability of voters to participate in an election, including by registering to vote, successfully voting, and having their ballot count after it has been voted.[40] Examples of historic vote denial instruments include "literacy tests, poll taxes, and registration barriers, all of which were notoriously common in the South prior to the enactment of the VRA in 1965."[41] Current vote denial cases generally concern voter identification laws, "limits on early and absentee voting, voter registration restrictions, and the rejection of provisional ballots."[42] Before it was gutted by *Shelby County*, Section 5 of the VRA effectively limited the number of vote denial cases litigated in the courts because the preclearance regime prevented discriminatory voting laws and practices from ever taking effect.[43]

In contrast to vote denial claims, vote dilution concerns political representation and influence.[44] Vote dilution involves limitations on a group's ability to elect candidates of their choice due to practice, policies, or procedures that unduly weaken political influence.[45] Common examples of vote dilution

35. South Carolina v. Katzenbach, 383 U.S. 301, 308 (1966).

36. *See* Allen v. State Bd. of Elections, 393 U.S. 544, 557 (1969) (finding an implied private right of action under Section 5 of the VRA). The Department of Justice also has the authority to enforce Section 2. 52 U.S.C. § 10308(d); *see also* U.S. DEPT. OF JUST., *supra* note 34.

37. 52 U.S.C. § 10301 (originally enacted as 42 U.S.C. § 1973 (1965)).

38. The Court in *Brnovich v. Democratic Nat'l Comm.* acknowledged that Section 2 applies to both vote denial and vote dilution and developed new guidelines applicable to vote denial claims. Brnovich v. Democratic Nat'l Comm., 141 S. Ct. 2321, 2338 (2021).

39. Daniel P. Tokaji, *Applying Section 2 to the New Vote Denial*, 50 HARV. C.R-C.L. REV. 439, 442 (2015) [hereinafter Tokaji, *Applying Section 2*].

40. *Id.*

41. *Id.*

42. *Id.*

43. *See The Effects of* Shelby County v. Holder, BRENNAN CTR. FOR JUST., (Aug. 6, 2018), https://www.brennancenter.org/our-work/policy-solutions/effects-shelby-county-v-holder [https://perma.cc/7Q5H-5RUR].

44. Daniel P. Tokaji, *The New Vote Denial: Where Election Reform Meets the Voting Rights Act,* 57 S.C. L. REV. 689, 691 (2006) [hereinafter Tokaji, *The New Vote Denial*]; *Applying Section 2*, *supra* note 39, at 442.

45. *See* Tokaji, *The New Vote Denial*, *supra* note 44, at 691; Tokaji, *Applying Section 2*, *supra* note 39, at 442.

include drawing districts in a manner that denies voters the ability to elect candidates of choice based on their race, color, or status as a language minority[46] and the use of at-large election systems.[47] While both vote denial and vote dilution claims can be brought under Section 2, this Article only addresses vote dilution and the legal framework that governs these claims.[48]

A. Legal Framework

Section 2(a) proscribes any "voting qualification or prerequisite to voting or standard, practice, or procedure . . . which results in a denial or abridgement of the right of any citizen of the United States to vote on account of race or color . . . ."[49] Section 2(b) explains factors that trigger subsection (a). Under subsection (b), a violation of Section 2 is established if, "based on the totality of circumstances," it is determined that the political process is "not equally open to participation" by members of a protected class, in that members of the protected class have less of an opportunity to "participate in the political process and to elect representatives of their choice."[50]

In *City of Mobile v. Bolden*, the Court determined that Section 2 required a finding of intentional discrimination, stating that "it is apparent that the language of § 2 no more than elaborates upon that of the Fifteenth Amendment,"[51] which is violated "only if motivated by a discriminatory purpose."[52] In response, Congress sought to enact a legislative override of the Court's decision. In 1982, Congress amended Section 2 of the VRA to permit plaintiffs to prove a Section 2 claim by proving that a policy has a discriminatory *effect*, whether or not that policy was adopted with an *intent* to discriminate.[53] These amendments created a new "results test,"[54] which prohibits electoral structures that "result[]" in a group of citizens "hav[ing] less opportunity than other members of the electorate to participate in

46. Tokaji, *Applying Section 2*, *supra* note 39; David P. Van Knapp, Annotation, *Diluting Effect of Minorities' Votes by Adoption of Particular Election Plan, or Gerrymandering of Election District, as Violation of Equal Protection Clause of Federal Constitution*, 27 A.L.R. Fed. 29 § 2(a) (1976).

47. *See* Pamela S. Karlan, *The Impact of the Voting Rights Act on African Americans: Second and Third-Generation Issues*, *in* VOTING RIGHTS AND REDISTRICTING IN THE UNITED STATES 121, 122 (Mark E. Rush ed., 1998); Tokaji, *Applying Section 2*, *supra* note 39, at 442.

48. The application of Section 2 to vote denial claims has been made more complicated by the Supreme Court's decision in *Brnovich v. Democratic Nat'l Comm.*, 141 S. Ct. 2321 (2021). The majority in *Brnovich*, however, did not address vote dilution and instead seemed to affirm current vote dilution jurisprudence. *Id.* at 2333.

49. 52 U.S.C. § 10301(a).

50. 52 U.S.C. § 10301(b).

51. 446 U.S. 55, 60 (1980).

52. *Id.* at 62.

53. S. REP. NO. 97-417, at 28 (1982); *see also* Thomas M. Boyd & Stephen J. Markman, *The 1982 Amendments to the Voting Rights Act: A Legislative History*, 40 WASH. & LEE L. REV. 1347, 1422–25 (1983).

54. *See* Christopher S. Elmendorf & Douglas M. Spencer, *Administering Section 2 of the Voting Rights Act After Shelby County*, 115 COLUM. L. REV. 2143, 2152 (2015).

the political process and to elect representatives of their choice" on account of race or color.[55] The amendment lacked a rigid standard of what it means for a racial minority to have unequal political opportunity. Instead, the revised Section 2 provides that an electoral structure is evaluated based on the "totality of circumstances."[56] The Senate Judiciary Committee report accompanying the passage of the 1982 Amendments to the Voting Rights Act recited a non-exhaustive list of factors to be weighed when making such an inquiry.[57]

After the 1982 Amendments, racially polarized voting became key to vote dilution claims. In 1986, in *Thornburg v. Gingles*, the Supreme Court clarified the evidentiary burden for Section 2's new language and set out the definitive test for racially polarized voting that is used in all Section 2 vote dilution cases.[58] Writing for a five-justice majority, Justice William Brennan determined that the plaintiffs had to satisfy certain preconditions in order to show that multimember districts operate to impair minority voters' ability to elect representatives of their choice.[59] Specifically, plaintiffs must show

> [f]irst, the minority group . . . is sufficiently large and geographically compact to constitute a majority in a single-member district. . . . Second, the minority group . . . is politically cohesive. . . . Third . . . that the white majority votes sufficiently as a bloc to enable it . . . usually to defeat the minority's preferred candidate.[60]

In *Bartlett v. Strickland*, the Court emphasized that in order to meet the first *Gingles* prerequisite, the party asserting a claim under Section 2 must show that the minority population in the district is greater than 50 percent.[61] The latter two conditions are the "racially polarized voting" requirement.[62] The *Gingles* preconditions serve two purposes: they are a limiting principle that could restrict

55. 52 U.S.C. § 10301(a)–(b).

56. 52 U.S.C. § 10301(b).

57. *See* S. REP. NO. 97-417, at 28–29 (1982) (listing "[t]ypical factors" that a plaintiff could show to establish a violation).

58. 478 U.S. 30, 48–51, 56–58, 62–63 (1986). Prior to *Gingles,* circuit and district courts had different definitions and standards for what constituted racially polarized voting. *See* Mary J. Kosterlitz, *Thornburg v. Gingles: The Supreme Court's New Test for Analyzing Minority Vote Dilution*, 36 CATH. U. L. REV. 531, 551 (1987). For example, in *Collins v. City of Norfolk*, 605 F. Supp. 377 (E.D. Va. 1984), a Virginia federal district court employed a three-factor test to identify racially polarized voting. The court agreed with defendants that three factors must be evaluated in determining whether racially polarized voting has occurred: (1) whether there is "white backlash," (2) "the voting patterns of black and white voters over a period of years," and (3) "whether whites attempt to limit the field of candidates." *Id.* at 386.

59. *Gingles*, 478 U.S. at 48–51.

60. *Id.* at 50–51.

61. 556 U.S. 1, 19–20 (2009).

62. Elmendorf, Quinn, & Abrajano, *supra* note 26, at 589. For more background on the legal scholarship that informed the *Gingles* majority, see James U. Blacksher, *From* Reynolds v. Sims *to* City of Mobile v. Bolden*: Have the White Suburbs Commandeered the Fifteenth Amendment?*, 34 HASTINGS L.J. 1, 43 (1982).

the number of cases that are heard, and they help courts determine whether the harm involved is a harm Section 2 of the VRA anticipated to remedy.[63] Only after these three *Gingles* preconditions are met will a court evaluate the totality of circumstances liability standard prescribed by the VRA.

After establishing the *Gingles* preconditions, courts then engage in a totality of the circumstances inquiry based on the factors outlined in the 1982 Senate Judiciary Committee Report.[64] The "Senate factors" considered as part of the totality of circumstances include (1) the extent of any history of official discrimination affecting minority voting rights in the jurisdiction at issue; (2) the extent to which voting in elections in the jurisdiction at issue is racially polarized; (3) the extent to which the jurisdiction has used malapportioned districts, majority vote requirements, or other voting procedures that enhance discrimination; (4) lack of minority access to any existing candidate slating processes; (5) the extent to which members of the protected class bear the effects of discrimination in areas such as education, employment, and health; (6) if there have been racial appeals, whether overt or subtle, used in political campaigns in the jurisdiction; and (7) the extent to which members of the protected class have been elected to office in the jurisdiction.[65]

In decisions after *Gingles*, the Supreme Court provided additional guidance on the conditions relevant to establishing a violation of Section 2.[66] In *Holder v. Hall*, the Court held that in addition to meeting the *Gingles* preconditions and satisfying the totality of circumstances test, courts must be able to "find a reasonable alternative practice as a benchmark against which to measure the existing voting practice."[67] Additionally, the Court found in *Abbott v. Perez* that a violation of Section 2 could not be established if the alternative to the allegedly dilutive district would not enhance minority representation.[68]

---

63. Elmendorf, Quinn, & Abrajano, *supra* note 26, at 589–90.

64. S. REP. NO. 97-417, at 28–29 (1982). The report states that the "factors are derived from the analytical framework used by the Supreme Court in *White* [*v. Regester*], as articulated in *Zimmer* [*v. McKeithen*]." *Id.* at 28 n.113. *See also* White v. Regester, 412 U.S. 755, 765–71 (1973); Zimmer v. McKeithen, 485 F.2d 1297, 1305 (5th Cir. 1973).

65. S. REP. NO. 97-417, at 28−29. The Senate Report lists two additional factors that some courts have considered: lack of responsiveness by elected officials to the needs of the protected class, and whether the policy underlying the state or jurisdiction's use of the voting qualification, practice, or standard is tenuous. *Id.* at 29. *See also* Thornburg v. Gingles, 478 U.S. 30, 44–45 (1986).

66. The Court has also clarified the types of claims to which Section 2 applies. In *Voinovich v. Quilter*, the Court held that Section 2 does not prohibit the creation of majority-minority districts unless these districts are necessary to remedy a statutory violation. Voinovich v. Quilter, 507 U.S. 146, 154–55 (1993).

67. 512 U.S. 874, 880 (1994). Finding that no objective, workable standard existed for choosing the size of a government body, the Court held that challenges to the size of a governing body are not cognizable under Section 2. *Id.* at 885.

68. 138 S. Ct. 2305, 2332 (2018) ("[I]t is hard to see how [the *Gingles*] standard could be met if the alternative to the districting decision at issue would not enhance the ability of minority voters to elect the candidates of their choice.").

In *Johnson v. De Grandy*, the Supreme Court discussed an additional consideration embedded in the Section 2 analysis: proportionality.[69] After satisfying the *Gingles* factors, during the totality of the circumstances considerations, courts may evaluate whether minority communities "form effective voting majorities in a number of districts roughly proportional to the minority voters' respective shares in the voting-age population."[70] The Court reasoned that Section 2 does not require the maximum number of possible majority-minority districts if the minority group already has a number of districts commensurate to its share of the voting age population. The Court held that "while proportionality is not dispositive in a challenge to single-member districting, it is a relevant fact in the totality of circumstances to be analyzed" in determining whether Section 2 has been violated.[71] In short, whether a minority community already has the opportunity to elect candidates proportionate to their population strength is one (not determinative) element of the totality of circumstances test.

In sum, the current Section 2 legal analysis first requires satisfying the three *Gingles* preconditions, including racially polarized voting, and then establishing that the totality of circumstances indicates a violation of Section 2. While "it will be only the very unusual case in which the plaintiffs can establish the existence of the three *Gingles* factors but still have failed to establish a violation of § 2 under the totality of circumstances,"[72] the additional totality of circumstances inquiry is important because evidence under the Senate Factors can demonstrate the impact that the vote dilution has had historically and contemporaneously on the minority group. The existence of racially polarized voting matters because when there is discrimination in the electoral sphere, that discrimination proliferates in all other aspects of life for the minority group in the jurisdiction when they are shut out or have their vote diluted.[73] As one scholar has noted, "racially polarized voting is not an aberration but a longstanding, pervasive, and continuing feature of

69. 512 U.S. 997 (1994).

70. *Id.* at 1000.

71. *Id.*

72. Jenkins v. Red Clay Consol. Sch. Dist. Bd. of Educ., 4 F.3d 1103, 1135 (3d Cir. 1993).

73. *See, e.g.*, V.O. KEY JR. & ALEXANDER HEARD, SOUTHERN POLITICS: IN STATE AND NATION (Caravelle ed., 1949) (discussing, in part, racism in southern politics and its influence on national politics and political procedures). *See also* EDWARD G. CARMINES & JAMES A. STIMSON, ISSUE EVOLUTION: RACE AND THE TRANSFORMATION OF AMERICAN POLITICS (1989) (investigating the development of race and racial equity as a defining theme across major elections and for party realignment); Michael W. Giles & Kaenan Hertz, *Racial Threat and Partisan Identification*, 88 AM. POL. SCI. REV. 317, 317–26 (1994) (testing the Power Theory, which asserts that where there is high threat posed by a minority group, the dominant group is expected to be more hostile politically); ROBERT HECKFELDT & CAROL WEITZEL KOHFELD, RACE AND THE DECLINE OF CLASS IN AMERICAN POLITICS (1989) (examining the role of racial polarization in the move of many white voters from the Democratic to the Republican party and the maintenance of racial cleavages in politics); Martin Gilens, Paul M. Sniderman & James H. Kuklinski, *Affirmative Action and the Politics of Realignment*, 28 BRITISH J. POL. SCI. 159, 159–83 (1998) (arguing that opposition to Affirmative Action encompasses the core of both political parties, despite the Democratic party's commitment to racial equity).

numerous jurisdictions' electoral histories," and an "overwhelming" body of evidence indicates that such patterns endure across the country.[74] As such, the Senate Factors provide context into the harms that are occurring in a jurisdiction and demonstrate why vote dilution is discriminatory.

B. Practical Implications

Redistricting and electoral schemes that do not comply with Section 2 can mean that minority voters are unable to elect candidates of their choice and are therefore shut out of the electoral process. Being shut out of the electoral process results in long-term harms. Indeed, vote dilution claims persist today just as they did four decades ago when the VRA was first signed into law. In 2006, the University of Michigan Law School Voting Rights Initiative researched judicial findings in 331 identified lawsuits under Section 2 of the VRA from 1982 to 2005.[75] Of the 331 lawsuits, courts found 92 documented violations of Section 2, and an additional 31 suits ended in a favorable determination for the plaintiff.[76] In total, only about 37% of plaintiffs succeeded in the lawsuits identified.[77] Section 2 has also been useful in litigating cases in jurisdictions covered by Section 5; plaintiffs bringing Section 2 cases in covered jurisdictions were more successful than those who brought cases in non-covered jurisdictions.[78] The study also found that from 1982 to 2005,

> African-American plaintiffs have brought the vast number of published claims (272 or 82.2%) under Section 2 since 1982, with an increasing number of cases involving Latino (97), Native American (12), and Asian American (7) plaintiffs. African-Americans were the sole plaintiffs in 93 (75.6%) of the successful decisions for plaintiffs. Of all lawsuits where any plaintiff achieved success, 16 involved multiple minority group plaintiffs. In addition, Latino plaintiffs won 7 lawsuits independently, and Native American plaintiffs won 5 published lawsuits.[79]

Section 2 has been instrumental in challenging discriminatory election systems and ensuring healthier democratic practices, and in some instances, voters who prevail on Section 2 claims are immediately impacted by increased

---

74. John M. Powers, *Statistical Evidence of Racially Polarized Voting in the Obama Elections, and Implications for Section 2 of the Voting Rights Act*, 102 GEO. L.J. 881, 891 (2013).

75. Ellen D. Katz, Margaret Aisenbrey, Anna Baldwin, Emma Cheuse, & Anna Weisbrodt, *Documenting Discrimination in Voting: Judicial Findings Under Section 2 of the Voting Rights Act Since 1982: Final Report of the Voting Rights Initiative, University of Michigan Law School*, 39 U. MICH. J. L. REFORM 643, 654 (2006).

76. *Id.* at 655.

77. *Id.*

78. *Id.*

79. *Id.* at 656.

representation. For example, in *Montes v. City of Yakima*, Latino[80] plaintiffs sued the City of Yakima, Washington for the use of an at-large election system to elect the seven seats on the Yakima City Council.[81] Yakima used a system in which Positions 1–4 on the city council had residency restrictions, while Positions 5–7 had no residency restriction.[82] Regardless of residency restrictions, however, during the general election all seats up for election were at-large races, meaning that the candidates running for each seat competed head-to-head, with all registered voters being able to cast one vote in each head-to-head race.[83] In *Montes*, the court found a Section 2 violation, meaning that plaintiffs satisfied the *Gingles* preconditions and that the totality of circumstances demonstrated that Yakima's electoral process was not equally open to participation by Latino voters.[84] Specifically, the court found that there was a history of voting-related discrimination,[85] racially polarized voting (with low white cross-over voting for Latino preferred candidates),[86] the presence of suspect voting practice or procedures that dilute Latino votes,[87] lingering effects of past discrimination,[88] and low success rates of Latino candidates in Yakima.[89] In the 37 years that the at-large system was used to elect the Yakima City Council, no Latino candidate had ever been elected, and the only Latina candidate to ever be appointed was ousted by a non-Latino challenger in the next election.[90] After finding a violation of Section 2 on summary judgment, the district court ordered the parties to confer

80. In this Article, we use the term Latino/Latina and not "Latinx" because the former is the term that a majority of community members use to identify themselves. *See* Luis Noe-Bustamante, Lauren Mora, & Mark Hugo Lopez, *About One-in-Four U.S. Hispanics Have Heard of Latinx, But Just 3% Use It*, PEW RSCH. CTR. (Aug. 11, 2020), https://www.pewresearch.org/hispanic/2020/08/11/about-one-in-four-u-s-hispanics-have-heard-of-latinx-but-just-3-use-it/ [https://perma.cc/5Q4G-8YLN].

81. Montes v. City of Yakima, 40 F. Supp. 3d 1377, 1385 (E.D. Wash. 2014).

82. *Id.* at 1385–86.

83. *Id.* at 1386.

84. *Id.* at 1414.

85. *Id.* at 1409–10.

86. *Id.* at 1410 ("Despite having received such strong support from Latino voters, the Latino candidate was defeated in every single race as a result of bloc voting by the non-Latino majority. In the dispositive elections, support for the Latino candidate (or Latino-preferred issue) among non-Latino voters ranged from **30.5%** (2009 City Council Position 7) to **42.6%** (2009 City Council Position 5). These low levels of 'crossover' support are highly indicative of majority bloc voting in this particular context; they demonstrate that, when presented with a choice between a Latino candidate and a non-Latino candidate, approximately 60% to 70% of non-Latino voters will vote for the non-Latino candidate.").

87. *Id.* at 1410–11.

88. *Id*. at 1413 (describing American Community Survey data demonstrating that Latinos are more likely to live below the poverty line, have lower median family incomes, have a lower rate of home ownership, have a lower rate of high school education, are less likely to have health insurance, and account for "for only 15% of City of Yakima employees, despite the fact that Latinos represent 33% of the City's working-age population").

89. *Id.* at 1413–14.

90. *Id.* at 1414.

and submit a new districting scheme that complies with Section 2.[91] In subsequent elections held under a new districting system, Latino voters were able to vote in the city's first three elected Latino city council members.[92] While Section 2 is not a panacea, it can be used to empower plaintiffs in their communities to meaningfully access the political process in their jurisdictions.

III.
PROVING RACIALLY POLARIZED VOTING

A. Introduction to Statistical Methods Used to Demonstrate Racially Polarized Voting

As discussed in Part II, racially polarized voting is a precondition for determining whether a Section 2 vote dilution claim is viable. Indeed, many courts, including the U.S. Supreme Court, have held that racially polarized voting is one of the most important—if not the most important—factors that must be proven in a vote dilution case.[93] The Eighth Circuit has similarly said that "the presence of racially polarized voting will ordinarily be the keystone of a vote dilution case."[94] The Fifth Circuit has also stated that "[e]vidence of racially polarized voting 'is

91. *Id.* at 1414–15.

92. Mike Faulk, *Four Years After Historic Wins for Latino Politicians, the Yakima City Council Is Getting Less Diverse*, CROSSCUT (Oct. 22, 2019), https://crosscut.com/2019/10/four-years-after-historic-wins-latino-politicians-yakima-city-council-getting-less-diverse [https://perma.cc/MNT9-CPL9]. Currently, a city manager appointed by the Yakima City Council manages the city's daily affairs. Lex Talamo, *Strong Mayor Killed: Yakima City Council Says They Won't Be the Ones to Take the Issue to the Voters,* YAKIMA HERALD (Feb. 5, 2020), https://www.yakimaherald.com/news/local/strong-mayor-killed-yakima-city-council-says-they-wont-be-the-ones-to-take-the/article_b9b8e511-3101-5007-a3b5-b23017ded4d0.html [https://perma.cc/QSF8-TXAT]. While Section 2 was important to changing the composition of the City Council, in 2020, members of the Yakima City Council attempted to pass a measure that would send a proposal to change the city's governance structure. *Id.* This means that the proposal to change the governance structure of the city would be voted on at-large by the still predominately white voting population of the city. This proposal would have stripped power from the City Council and replaced the city manager with a mayor elected in a citywide vote. *Id.* This "strong mayor" form of government would dilute the vote of Latinos. The council initially approved the measure 4-3 but changed course when two lawsuits were filed in response. *Id.* One of these lawsuits was filed by the UCLA Voting Rights Project on behalf of Latino voters in the city after the Project was contacted by Latino voters and Columbia Legal Services, a legal aid organization in Eastern Washington. *Id.*

93. *See, e.g.*, Thornburg v. Gingles, 478 U.S. 30, 67 (1986) (calling racially polarized voting "one of the most important elements of a vote dilution claim"); Citizens for a Better Gretna v. City of Gretna, 834 F.2d 496, 499 (5th Cir. 1987) (calling racial bloc voting the "linchpin" of a vote dilution claim); Lucas v. Townsend, 967 F.2d 549, 551 (11th Cir. 1992) ("Absent a showing of racially polarized voting, the challenged practice does not affect minority voting rights and cannot cause a discriminatory result."); Terrebonne Parish Branch NAACP v. Jindal, 274 F. Supp. 395, 410 (M.D. La. 2017) (stating that the two most important Senate Factors are "the existence of racially polarized voting and the extent to which minorities are elected to public office" (citing Clark v. Calhoun, 88 F.3d 1393, 1397 (5th Cir. 1996))).

94. Buckanaga v. Sisseton Indep. Sch. Dist., 804 F.2d 469, 473 (8th Cir. 1986).

the linchpin of a section [sic] 2 vote dilution claim.'"[95] As of 2005, most plaintiffs who satisfied the *Gingles* preconditions obtained a "favorable outcome."[96]

The focus of the racially polarized voting inquiry is (1) whether the minority group is "politically cohesive" and (2) whether the majority is voting as a bloc against minority-preferred candidates, thereby preventing minority voters from electing candidates of their choice.[97] Since *Thornburg v. Gingle*s, plaintiffs have had to satisfy both conditions, among others.[98] The test serves three main purposes. First, the test is thought of as providing some clear and objective standards for a finding of racial vote dilution.[99] Second, the preconditions can be viewed as a limiting principle that filter strong claims from weaker claims and restrict the number of cases that require sensitive decisions about "racial fairness in the distribution of political opportunity."[100] Finally, the preconditions enable the court to efficiently make a "normative diagnosis" about whether the harm in question might rise to the nature of the harm contemplated by Section 2.[101]

The racially polarized voting burden can be a substantial hurdle for plaintiffs to prove in vote dilution cases involving a jurisdiction with limited data on voter behavior. For starters, lower courts have not established a consistent quantitative standard for what is or is not legally significant racially polarized voting.[102] For example, the United States District Court for the Northern District of Ohio stated in *United States v. City of Euclid* that

> while courts have found certain particular statistical or mathematical outcomes to be compelling evidence *in the context of the cases before them,* no decision out of either the Supreme Court or the Sixth Circuit (or any other Circuit for that matter) requires the use of a particular statistical methodology, or demands a particular statistical outcome before a court may conclude that racial bloc voting exists.[103]

The degree of legally significant minority cohesiveness and the level of white bloc voting sufficient to defeat a minority preferred candidate depends on a variety

95. Westwego Citizens for Better Gov't v. City of Westwego, 946 F.2d 1109, 1118 (5th Cir. 1991) (alteration in original).

96. *See* Katz, Aisenbrey, Baldwin, Cheuse, & Weisbrodt, s*upra* note 75, at 660.

97. *Gingles*, 478 U.S. at 56.

98. *See supra* Part II.

99. *See* Elmendorf, Quinn, & Abrajano, *supra* note 26, at 603. Additionally, a finding of racially polarized voting can determine the possibility of a judicially mandated remedy should the court find in favor of liability (for example, by ensuring that different voting district boundaries would actually grant a minority group the political opportunity demanded by Section 2).

100. *Id.* at 589, 598–99.

101. *Id.* at 589−90.

102. *See* Jenkins v. Red Clay Consol. Sch. Dist. Bd., 4 F.3d 1103, 1126 (1993) (holding that the third *Gingles* precondition "may be satisfied with a variety of evidence, including lay testimony or statistical analyses of voting patterns").

103. 580 F. Supp. 2d 584, 596 (N.D. Ohio 2008).

of factual circumstances, meaning the test hardly lends itself to bright-line rules.[104] Beyond uncertainty as to *how much* evidence needs to be presented, there is also uncertainty about *what kind* of evidence is required. As noted above, the courts have never articulated a requirement as to the type of data that a plaintiff can depend upon or the kind of statistical models that can be used to interpret this data[105]—the uniqueness of each case necessarily requires a flexible approach. While both qualitative and quantitative information is often considered,[106] the courts have articulated a preference for certain types of quantitative data over others, including qualitative. While the test must remain flexible in the face of differing circumstances, plaintiffs often cannot predict the type of evidence a fact-finder might find most persuasive. This suggests that new and better social science methods should be offered to bolster voting rights plaintiffs' cases.

Historically, the process of showing racially polarized voting has depended on making inferences via statistical models using voter turnout statistics, demographic data, and election results in the relevant jurisdiction.[107] Individual-level vote choice, which in large statewide elections may be captured through exit polls, is often unknown in local elections.[108] Given the ubiquity of the secret ballot format,[109] specific information about which candidate(s) each individual voter actually cast a ballot for is not publicly available.[110] The only information

104. *See* Cottier v. City of Martin, No. 02-5021-KES, 2005 WL 6949764, at *22 (W.D.S.D. Mar. 22, 2005) ("[N]o mathematical formula or simple doctrinal test is available to determine whether plaintiffs satisfied the third factor.").

105. *See, e.g.*, Rodriguez v. Harris County, 964 F. Supp. 2d 686, 757–58 (S.D. Tex. 2013), *aff'd sub nom.* Gonzalez v. Harris County, 601 F. App'x 255 (5th Cir. 2015) ("Courts have relied on various statistical methods." (*See, e.g.,* Houston [v. Lafayette County], 56 F.3d [606,] 611 [(5th Cir. 1995)] (use of bivariate ecological regression and extreme case analysis); Clark [v. Calhoun County], 88 F.3d [1393,] 1397 [(5th Cir. 1996)] (expert employed regression and homogeneous precinct analysis); Benavidez [v. City of Irving], 638 F. Supp. 2d [709,] 723–24 [(N.D. Tex. 2009)] (ecological inference).").

106. *See* Cottier v. City of Martin, 445 F.3d 1113, 1118 (8th Cir. 2006) ("Proving political cohesiveness requires evaluating elections through statistical and non-statistical evidence."); *see also* Sanchez v. Bond, 875 F.2d 1488, 1494 (10th Cir. 1989) ("The experiences and observations of individuals involved in the political process are clearly relevant to the question of whether the minority group is politically cohesive. This testimony would seem to be required if the court is to identify the presence or absence of distinctive minority group interests.").

107. Klein, Sacks, & Freedman, *supra* note 30, at 396.

108. *See* Eric A. Hanushek & John M. Quigley, *Citizen Turnout and Self-Interested Voting: Inferring Preferences from Secret Ballots* 1 (Inst. of Bus. & Econ. Rsch., Working Paper No. 84-79, 1981).

109. *See* James D'Angelo & Brent Ranalli, *How the Secret Ballot Ended the Gilded Age*, CONG. RSCH. INST. (Aug. 9, 2020), https://www.congressionalresearch.org/SecretBallot.html [https://perma.cc/A6TU-THX8] ("Beginning in the 1830s, the iconic symbol of democracy . . . was the transparent, glass globe, ballot box. But, in the late 1800s with partisanship, campaign finance and inequality soaring to all-time highs, the United States began to roll out the secret ballot."); Burson v. Freeman, 504 U.S. 191, 206 (1992) ("After an unsuccessful experiment with an unofficial ballot system, all 50 States, together with numerous other Western democracies, settled on the same solution: a secret ballot secured in part by a restricted zone around the voting compartments.").

110. James D. Greiner & Kevin M. Quinn, *Exit Polling and Racial Bloc Voting: Combining Individual-Level and RxC Ecological Data*, 4 ANNALS APPLIED STAT. 1774, 1774 (2010).

available about an individual voter's behavior in a given election is whether or not they voted.[111] At an aggregate level, voting patterns, where they can be discovered, can be aligned with demographic information.

The main source of both voting age population data and citizenship data by race published by the Census Bureau now comes from the American Community Survey ("ACS").[112] The ACS is an annual, nationwide survey that collects demographic information, including age, race, ethnicity, and citizenship, from a sample of roughly two to three million households.[113] With this data, the Census Bureau is then able to estimate both the voting age population ("VAP") and citizen, voting-age population ("CVAP") of states, counties, census tracts, and census block groups.[114] VAP is also available from the decennial Census's PL-94-171 file and is used for redistricting. At the state level, the Census Bureau also implements the Current Population Survey ("CPS") monthly for labor force statistics, and every two years implements a November supplement for voting and registration data.[115] The total sample size of the CPS and supplemental data nationwide is only about 60,000 households, which allows for statistically sound state-level estimates, but can present sample size challenges for smaller jurisdictions.[116] In contrast, the ACS typically contains interviews of approximately 2 million households. The Census Bureau aggregates ACS data over five-year periods in order to provide more estimates for small areas, such as census tracts and block groups.[117] Currently, about 3–3.5 million households are

111. *See* Drew DeSilver, *Voter Files: What Are They, How Are They Used and Are They Accurate*, PEW RSCH. CTR. (Feb. 15, 2018), https://www.pewresearch.org/fact-tank/2018/02/15/voter-files-study-qa/ [https://perma.cc/NS23-6CWX].

112. *See About the ACS*, U.S. CENSUS BUREAU (Jan. 14, 2021), https://www.census.gov/programs-surveys/acs/about.html [https://perma.cc/ETT5-DVSU]; *Citizenship*, CENSUS REP. (Aug. 29, 2021), https://censusreporter.org/topics/citizenship/ [https://perma.cc/P66S-G4VR]; *see generally American Community Survey Information Guide*, U.S. CENSUS BUREAU (Oct. 2017), https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf [https://perma.cc/GSH7-AY8N] (providing overview of ACS survey and data collection).

113. *See American Community Survey Sample Size*, U.S. CENSUS BUREAU (last visited Dec. 18, 2021), https://www.census.gov/acs/www/methodology/sample-size-and-data-quality/sample-size [https://perma.cc/QT4Y-96HD]; *American Community Survey Information Guide*, *supra* note 112, at 6, 10.

114. *See American Community Survey Information Guide*, *supra* note 112, at 11; *Citizenship*, *supra* note 112.

115. *See Current Population Survey*, NOVEMBER 2020 VOTING AND REGISTRATION SUPPLEMENT, U.S. CENSUS BUREAU 2-2, 3-1 (Nov. 2020), https://www2.census.gov/programs-surveys/cps/techdocs/cpsnov20.pdf [https://perma.cc/E4XP-6AM9].

116. *See Methodology*, U.S. CENSUS BUREAU (Oct. 8, 2021), https://www.census.gov/programs-surveys/cps/technical-documentation/methodology.html [https://perma.cc/M2CZ-J5RE].

117. *See* U.S. CENSUS BUREAU, A COMPASS FOR UNDERSTANDING AND USING AMERICAN COMMUNITY SURVEY DATA: WHAT GENERAL DATA USERS NEED TO KNOW 24 (2009), https://www.census.gov/content/dam/Census/library/publications/2009/acs/ACSResearch.pdf [https://perma.cc/EY27-G56T].

sampled each year, out of more than 140 million households.[118] The five-year data represent a sample of roughly 10% of all households (each single year ACS successfully interviews roughly 2% of all households), from which inferences are made to the population, given published margins of error.[119] This information on the eligible voting population can be matched with the actual precinct-level vote results in a given election. With this information, the voting patterns of majority and minority voters can be extrapolated using statistical methods called ecological inference. Scholars and practitioners, however, should take note that CVAP data represents racial population estimates for all *eligible* voters in or around the voting precinct, not the racial population estimates for the *actual* voters in an election.[120]

Several ecological methods are available to assess the *Gingles* preconditions of minority cohesion and white bloc voting.[121] Ecological Inference (EI) "has been the benchmark method courts use in evaluating racial polarization in voting rights lawsuits, and has been used widely in comparative politics research on group and ethnic voting patterns."[122] Two variations of EI that have emerged are referred to as King's EI and EI: RxC.[123] The two methods are closely related, and Professor Gary King, the creator of King's EI,[124] was a co-author and collaborator on the RxC method.[125] Generally speaking, both methods take ecological data in the aggregate—such as precinct vote totals and racial demographics—and use Bayesian statistical methods to find voting patterns by regressing candidate choice against racial demographics within the aggregate precinct.[126] King's EI is sometimes referred to as the iterative approach, in that it runs an analysis of each candidate and each racial group in iterations,[127] whereas the RxC method allows multiple candidates and multiple racial groups to be estimated simultaneously in

118. *See American Community Survey Sample Size*, *supra* note 113; *American Community Survey Information Guide*, *supra* note 112, at 10.

119. JORGE CHAPA, ANA HENDERSON, AGGIE JOOYOON NOAH, WERNER SCHINKIV, & ROBERT KENGLE, CHIEF JUST. EARL WARREN INST. ON L. & SOC. POL'Y, REDISTRICTING: ESTIMATING CITIZEN VOTING AGE POPULATION 3–4 (2011), https://www.law.berkeley.edu/files/Redistricting_PolicyBrief4_forWeb.pdf [https://perma.cc/75U8-RMFT] (explaining use of the Census (ACS data) to develop CVAP estimates, as well as limitations and methods for accurate estimation); *see American Community Survey Sample Size*, *supra* note 113; *American Community Survey Information Guide*, *supra* note 112, at 10.

120. *See* CHAPA, HENDERSON, NOAH, SCHINKIV, & KENGLE, *supra* note 119 at 4.

121. For an approachable overview of this material, see BRUCE M. CLARKE & ROBERT TIMOTHY REAGAN, REDISTRICTING LITIGATION: AN OVERVIEW OF LEGAL, STATISTICAL, AND CASE-MANAGEMENT ISSUES 53–64 (2002).

122. Collingwood, Oskooii, Garcia-Rios, & Barreto, *supra* note 28, at 93.

123. *See* Barreto, Collingwood, Garcia-Rios, & Oskooii, *supra* note 3, at 271.

124. *See generally* GARY KING, A SOLUTION TO THE ECOLOGICAL INFERENCE PROBLEM RECONSTRUCTING INDIVIDUAL BEHAVIOR FROM AGGREGATE DATA (1997) (presenting King's EI method).

125. *See* Ori Rosen, Wenxin Jiang, Gary King, & Martin Tanner, *Bayesian and Frequentist Inference for Ecological Inference: Case*, 55 STATISTICA NEERLANDICA 134, 134–46 (2001).

126. Grofman & Merrill, *supra* note 3, at 123–24; *see generally* Barreto, Collingwood, Garcia-Rios, & Oskooii, *supra* note 3 (describing and comparing two EI methods).

127. *See* Barreto, Collingwood, Garcia-Rios, & Oskooii, *supra* note 3, at 3.

one model.[128] In essence, both versions of EI operate as described above: by compiling data on the percentage of each racial group in a precinct and merging that with precinct-level vote choice from relevant election results.[129] One popular software program *eiCompare*, imports data and runs both King's EI and RxC models and offers comparison diagnostics.[130] Researchers have concluded that both EI and RxC tend to produce similarly reliable regression estimates of vote choice.[131]

The EI models are agnostic on what type of input data political scientists use for racial demographics. It can be CVAP data on *eligible* voters, it can be a Spanish surname analysis of *registered* voters, or it can be a BISG estimate of race of *actual* voters. The models will perform the same analysis and produce inferences about voter preference by race.[132] One team of political scientists and demographers recommends using data of people who voted first, where possible and then perhaps those registered, before moving to less precise estimates like all eligible voters (CVAP).[133] BISG is one tool researchers can then use to understand the race and ethnicity of voters in a given precinct before conducting ecological inference.

The chart below provides a brief summary of the different types of data that may be inputted into the two leading ecological inference models to produce estimates on voting patterns by race.

Table 1: Summary of Race Input Data that Can Be Inputted Into EI Models

CVAP (ACS)	Surname list	BISG
Census 5-year average ACS data for *eligible* voters within a block group, matches that to a voting precinct.	Census surname list to match to voter file of actual voters, but typically only for Latino or Asian voters, within a precinct.	Census surnames for all racial groups, combined with Census Block racial demographics for actual voters within a precinct.

B. Potential Limitations of Census Data in EI Models

The use of ACS's CVAP data for racially polarized voting analysis at the local level has some underlying methodological limitations that render it incomplete in some contexts. Still, it is generally an accurate measure of voters' race and ethnicity. As social scientists, our aim is to work with existing tools, but also sharpen those tools to help improve our analyses in the future. While CVAP data

128. *Id.*
129. *Id.* at 28.
130. Collingwood, Oskooii, Garcia-Rios, & Barreto, *supra* note 28, at 92, 94.
131. *Id.* at 93–94.
132. *See* Barreto, Collingwood, Garcia-Rios, & Oskooii, *supra* note 3 (describing the performance of King's El and El: RxC with diverse data).
133. *See* Hood, Morrison, & Bryan *supra* note 26, at 547 ("Given a choice, the order of preference for data type would be turnout, otherwise registration, otherwise CVAP or VAP.").

is often useful for estimating districtwide population, it has three possible limitations when used at the much smaller precinct level. First, 5-year CVAP data is roughly a 10% household sample, which is limited compared to the 100% household sample of the decennial Census.[134] Additionally, a five-year estimate of CVAP information is somewhat dated, as the most current ACS data that is available includes some past portion of the sample based on past responses because it is from the past five years. All CVAP estimates therefore contain an upper/lower confidence interval or margin of error.[135] Second, CVAP data does not account for who the actual voters are—rather, it reflects sample data on the overall pool of eligible adult citizens.[136] Third, CVAP data is available at the census block group level, which does not align neatly to local precinct boundaries, often creating a mismatch.[137] Each of these issues could introduce some amount of noise and uncertainty into the precinct-level race estimates produced by CVAP. Taken together, these issues can sometimes result in less accurate estimates of race or ethnicity of a voter, especially in areas with lower turnout rates or a small number of voting precincts.[138] If turnout is unequal across racial groups, data on *all eligible voters* could be misleading as to the racial demographics of the *actual voters* in each precinct. In some instances, however, CVAP racial data aggregated

134. Note that in some jurisdictions where there are not large immigrant populations, VAP data might be used instead of CVAP. If VAP data comes from the decennial Census, it can avoid the limitation of only consisting of a 10% sample as the decennial Census comes close to a 100% household count. Decennial Census VAP is often quite accurate in equal turnout jurisdictions for African American or Native American voters. However, VAP is like CVAP in that it is a count of potentially eligible voters, not actual voters, and that its census geography is not necessarily the same as a voting precinct geography. *See 2020 Census Nonresponse Followup Rates Available at Local Level*, U.S. CENSUS BUREAU (Sept. 3, 2020), https://www.census.gov/newsroom/press-releases/2020/nonresponse-rates-local-level.html [https://perma.cc/43RZ-YALE]; Pat Cantwell, *How We Complete the Census When Households or Group Quarters Don't Respond*, U.S. CENSUS BUREAU (Apr. 16, 2021), https://www.census.gov/newsroom/blogs/random-samplings/2021/04/imputation-when-households-or-group-quarters-dont-respond.html [https://perma.cc/73MT-BN3K]; *see also* CHAPA, HENDERSON, NOAH, SCHINKIV, & KENGLE, *supra* note 119, at 3–4; *American Community Survey Sample Size*, *supra* note 113; *American Community Survey Information Guide*, *supra* note 112, at 10.

135. CHAPA, HENDERSON, NOAH, SCHINKIV, & KENGLE, *supra* note 119, at 4.

136. *See id.*

137. *See infra* Figure 2; CHAPA, HENDERSON, NOAH, SCHINKIV, & KENGLE, *supra* note 119, at 5.

138. This is because when there are a small number of voting precincts, the conservative estimates of CVAP within each Census block are aggregated together, often leading to an underestimation of voters within a given precinct, as well as geographical mismatches between precinct and block boundaries. *See* CHAPA, HENDERSON, NOAH, SCHINKIV, & KENGLE, *supra* note 119, at 3–4; *American Community Survey Sample Size*, *supra* note 113 (noting that CVAP estimates for census blocks are often less accurate as a result of the estimates being drawn from small population samples aggregated over multiple years often leading to high margins of error, potential underestimation, and data suppression in some precincts); *American Community Survey Information Guide*, *supra* note 112, at 10; Jesse T. Clark, John A. Curiel, & Tyler S. Steelman, *Minmaxing of Bayesian Improved Surname Geocoding and Geography Level Ups in Predicting Race*, POL. ANALYSIS 1–7 (2021).

to precincts is the *only* available data in particular jurisdictions.[139] Despite its potential limitations, in many circumstances CVAP can be used successfully to conduct analysis to support a finding of racially polarized voting. We offer these limitations as a cautionary note that analysts and litigators should consider when deciding what type of demographic data to use in their analyses.

The first consideration is the 10% sample on which the five-year ACS CVAP data relies. The ACS sample might have few, or even zero, responses from smaller census block groups. Even when the ACS data are aggregated over a five-year period to increase the number of observations, census block groups often report large margins of error.[140] After all, the ACS only samples approximately 3.5 million households each year.[141] Consider recent data from North Carolina as an example. Population estimates by race from the ACS 2013–2017 five-year aggregate data for Durham County, North Carolina illustrate the margin of error issue at census block group geography.[142] Figure 1 is a screenshot from the U.S. Census website that shows, for example, that Census Tract 2, Block Group 2 the white non-Hispanic population is estimated to be 395 people, with an accompanying range of plus or minus 175 people.[143] Likewise, there are an estimated 414 Black people, +/- 168 and an estimated 167 Hispanic or Latino people, +/- 186.[144]

139. *See id.*

140. U.S. CENSUS BUREAU, UNDERSTANDING AND USING ACS SINGLE-YEAR AND MULTIYEAR ESTIMATES 13 (2018), https://www.census.gov/content/dam/Census/library/publications/2018/acs/acs_general_handbook_2018_ch03.pdf [https://perma.cc/WP65-MT2T] (noting that an ACS five-year estimate for child poverty statistics in Rice County, Minnesota had a confidence interval spanning six percentage points, compared to a confidence interval of eight percentage points for the one-year estimate).

141. CHAPA, HENDERSON, NOAH, SCHINKIV, & KENGLE, *supra* note 119, at 3.

142. *See Durham County, North Carolina, Hispanic or Latino Origin by Race*, U.S. CENSUS BUREAU, https://data.census.gov/cedsci/table?q=B03002%3A%20HISPANIC%20OR%20LATINO%20ORIGIN%20BY%20RACE&g=1400000US37063000101%241500000,37063000102%241500000,37063000200%241500000&tid=ACSDT5Y2017.B03002 [https://perma.cc/FA9B-ZAEM] (last visited Feb. 21, 2022) (linking to underlying data from US Census Bureau table "B03002: HISPANIC OR LATINO ORIGIN BY RACE, 2017: ACS 5-Year Estimates Detailed Tables").

143. *Id.*

144. *Id.*

Figure 1: Example of Census ACS Published Margin of Error for Select Census Block Group in Durham County, N.C.

	Block Group 2, Census Tract 2, Durham County, North Carolina	
Label	Estimate	Margin of Error
Total:	1,027	±339
Not Hispanic or Latino:	860	±273
White alone	395	±175
Black or African American alone	414	±168
American Indian and Alaska Native alone	0	±12
Asian alone	0	±12
Native Hawaiian and Other Pacific Islander alone	0	±12
Some other race alone	50	±72
Two or more races:	1	±2
Two races including Some other race	0	±12
Two races excluding Some other race, and three or more races	1	±2
Hispanic or Latino:	167	±186
White alone	112	±169
Black or African American alone	0	±12
American Indian and Alaska Native alone	0	±12
Asian alone	0	±12
Native Hawaiian and Other Pacific Islander alone	0	±12
Some other race alone	54	±84

Table 2 compares the racial population estimates of block groups, depending on if the mid-point, lower, or upper bound of the estimate used for select block groups.[145] Across seven block groups in Durham that we randomly selected, population counts vary widely when the margin of error is taken into account. The same Block Group (Tract 2, Block Group 2) could be anywhere from 21% to 56% white non-Hispanic, while the Black population might make up anywhere from 24% to 57%, and the Hispanic and/or Latino[146] population from 0% to 34%. While the mid-point of the estimate is likely the closest to reality,[147] the small sample size of the ACS reveals that census race estimates can contain considerable noise and uncertainty at the block group level that must be carefully considered by the racially polarized voting analyst. In short, this census block group *could be* 25% white and 75% minority; or it *could be* 55% white and 45% minority—two very different racial compositions that we would be feeding into our EI model.

145. *See id.*

146. *Id.* The Census categorizes Hispanic or Latino as a single ethnic category, separate from race, and often uses the terms Hispanic and Latino interchangeably.

147. For more on central limit theorem and why the mid-point of the estimate is closest to the mean, see generally *Central Limit Theorem,* B.U. SCH. PUB. HEALTH, https://sphweb.bumc.bu.edu/otlt/mph-modules/bs/bs704_probability/BS704_Probability12.html [https://perma.cc/KEE6-Q2HK] (last visited Mar. 6, 2022).

Table 2: 2013−2017 ACS 5-year Racial Population Data for Selected Census Block Groups in Durham County, N.C.

	Tract 1.01, BG 2			**Tract 1.02, BG 2**			**Tract 2.0, BG 1**			**Tract 2.0, BG 2**		
	low	est	up	low	est	up	low	est	up	low	est	up
White	197	300	403	607	830	1053	203	337	471	220	395	570
Black	658	948	1238	167	311	455	256	476	696	246	414	582
Asian	-12	0	12	16	66	116	-12	0	12	-12	0	12
Latino	96	433	770	143	329	515	-66	483	1032	-19	167	353
Total	1705			1620			1320			1027		
	%	%	%	%	%	%	%	%	%	%	%	%
White	12%	18%	24%	37%	51%	65%	15%	26%	36%	21%	38%	56%
Black	39%	56%	73%	10%	19%	28%	19%	36%	53%	24%	40%	57%
Asian	-1%	0%	1%	1%	4%	7%	-1%	0%	1%	-1%	0%	1%
Latino	6%	25%	45%	9%	20%	32%	-5%	37%	78%	-2%	16%	34%

low = lower bound of the estimate; est = midpoint of the estimate; up = upper bound of the estimate

Some defendants have asserted—unsuccessfully—that uncertainty in ACS estimates calls the reliability of racially polarized voting analysis into question. Voting rights defendants in Texas have argued this point in court, citing "high margins of error for the ACS data" and arguing that "combining data from the ACS and Census is statistically problematic," and that "there are various errors and uncertainties in estimating the number, location, and citizenship status" of minority voters like Hispanics.[148] The district court, however, rejected defendants' arguments, stating, "the five-year ACS is the most reliable version of the ACS for analyzing small populations" and that "the ACS is the only source of citizenship data collected by the Census Bureau."[149] Therefore, while its limitations should be acknowledged and potential improvements (such as BISG) explored, voting rights advocates should feel confident using ACS CVAP data in Section 2 vote dilution litigation.

A second limitation of ACS CVAP data is that it does not represent the actual voters who participated in the election for which voting patterns are being analyzed. CVAP represents the totality of all eligible voters, including those not

148. Fabela v. City of Farmers Branch, No. 3:10-cv-1425, 2012 WL 3135545, at *7 (N.D. Tex. Aug. 2, 2012); *see also* Cisneros v. Pasadena Indep. Sch. Dist., No. 4:12-cv-2579, 2014 WL 1668500, at *9 (S.D. Tex. Apr. 25, 2014) (dismissing plaintiff's arguments about the insufficiency of ACS CVAP data).

149. *Fabela*, 2012 WL 3135545, at *7.

registered to vote, who reside in or around the precinct.[150] Durham, North Carolina, can offer an example once again. According to the ACS, the city of Durham has an adult voter-eligible voter population of 186,727.[151] In the November 2019 election for mayor of Durham, a total of 34,867 votes were cast, or 19% of the total CVAP.[152] When we plug CVAP data into our EI models to predict vote choice, we therefore have a less precise starting point for the population, especially in the case of local elections with lower turnout rates. In this case, 81% of the citizen voting-age population did not vote, yet they are included in the ACS CVAP data.[153] These limitations can be mitigated, and ACS CVAP data has frequently been accepted by courts,[154] but using the actual voter rolls for people who voted in the election being analyzed should always be the standard when such data are available.

Third, ACS CVAP data can result in misalignment between voting precinct and census block group, misplacing voters and altering the data for individual precincts (see Figure 2 below). When interested in people in a specific precinct, it is best to gather data only on voters who live within that precinct's boundaries. As election data are provided at the level of individual precinct—and not Census block group—precinct-level data is necessary to conduct EI. However, CVAP data is reported at the block group level, not at the voting precinct level. As such, there is always some degree of geographic misalignment between census boundaries and precinct boundaries, and demographers have to split census block groups and try to assign them to a voting precinct to get an accurate picture of the racial makeup of the precinct. As depicted in Figure 2,[155] for an actual voting precinct (#19) in Durham, North Carolina, as many as nine different census block groups partially overlap with a precinct. Some are entirely inside the precinct, others are half-in and half-out, and others only have a small portion inside the precinct. This

150. Census ACS does not report data at the level of the voting precinct. The smallest unit the ACS uses is census block group. *See* U.S. CENSUS BUREAU, GEOGRAPHIC AREAS COVERED IN THE ACS, https://www.census.gov/content/dam/Census/library/publications/2020/acs/acs_geography_handbook_2020_ch01.pdf [https://perma.cc/2FWJ-KGPU].

151. *Durham County, North Carolina, Sex by Age by Nativity and Citizenship Status*, U.S. CENSUS BUREAU, https://data.census.gov/cedsci/table?g=1600000US3719000&tid=ACSDP5Y2019.DP05 [https://perma.cc/5MYT-3664] (last visited Mar. 6, 2022) (linking to underlying data from US Census Bureau table "B05003: SEX BY AGE BY NATIVITY AND CITIZENSHIP STATUS, 2019: ACS 5-Year Estimates Detailed Tables").

152. *11/05/2019 Official Municipal Election Results—Durham*, N.C. STATE BD. ELECTIONS, https://er.ncsbe.gov/?election_dt=11/05/2019&county_id=32&office=ALL&contest=0 [https://perma.cc/SD8F-LBDQ] (last visited Oct. 14, 2021).

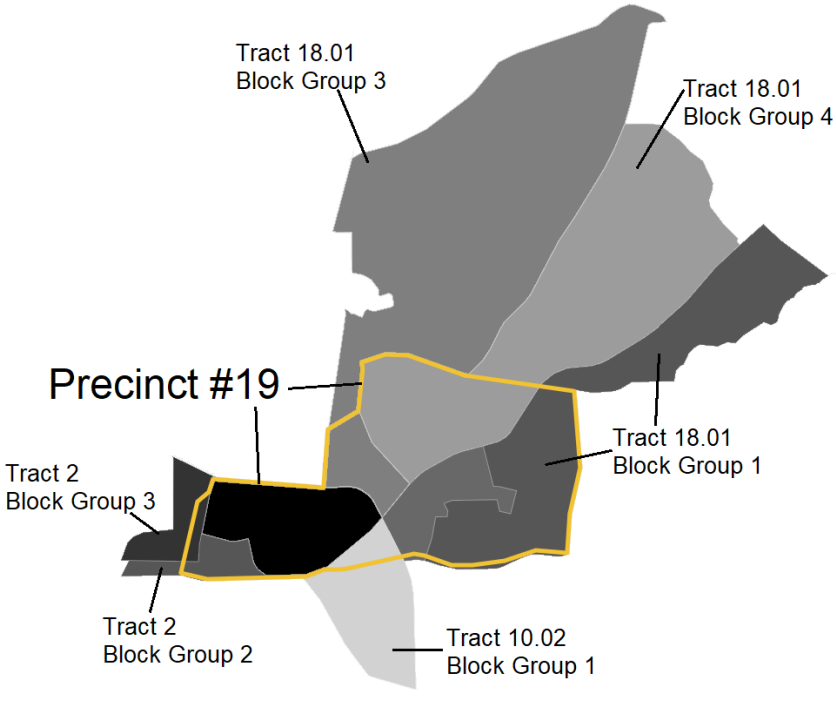
153. Incorporating a separate candidate column in the underlying dataset for "no vote" may reduce the error associated with using CVAP as racial data input. This could be in either iterative EI or RxC as an independent outcome variable. However, this does not solve precinct-level race by turnout issues. In fact, the model using CVAP must still estimate who did and did not vote within a precinct by race, adding an additional layer of uncertainty.

154. *See, e.g.*, sources cited *supra* notes 148–49.

155. Data visualization created by Authors and on file with Authors. Map and data come from U.S. Census American Community Survey (ACS) Block Group boundaries and Durham County Voting Precinct Map.

misalignment may reduce the precision of using census block groups to estimate precinct voters' race.

Figure 2: Misalignment Between Census Block Group and Voting Precinct in Durham, N.C.



Despite these limitations, EI using ACS CVAP data can, and should, be reliably used to provide vote choice estimates. When working with the decennial Census data, social scientists might rely on VAP data, especially in areas that have larger Black or white, non-Hispanic populations. These data have been proven time and again to produce reliable results, especially in jurisdictions with a larger number of precincts. Our critique above is intended to point out some limitations with CVAP data and open the door for scholars to work with more precise voter data using BISG. In some jurisdictions, access to electronic lists of actual voters is not readily available.[156] In jurisdictions with a larger total number of precincts, and greater variation in racial populations across precincts, EI using VAP or CVAP might produce vote choice estimates similar to those with EI using data on

156. *See Access to and Use of Voter Registration Lists*, NCSL (Jan. 3, 2022), https://www.ncsl.org/research/elections-and-campaigns/access-to-and-use-of-voter-registration-lists.aspx [https://perma.cc/2ZWH-VETU].

actual voters.[157] When other methods are available, like those described here, they should be considered instead. Nevertheless, EI with VAP or CVAP data, standing alone, is enough to establish racially polarized voting in the vast majority of political jurisdictions and has been accepted by courts as such. The data that CVAP compiles on race within a voting precinct is still likely to be correlated with the true racial demographics of those who actually voted, and thus, it is still likely to produce a reasonable coefficient estimate of vote choice.[158] The issue is that social scientists wish to reduce error as much as possible, and as we have outlined, certain limitations inherent within CVAP provide an opportunity for error to creep into the estimates. This is where BISG can help. BISG offers an opportunity to obtain more precise estimates of the racial and ethnic demographics of actual voters to use as the input variable in a model predicting how they actually voted. This should help us get closer to the most accurate estimate of vote choice using ecological data and provide even more precise racially polarized voting analysis.

ACS CVAP data—or VAP from the decennial Census, when appropriate—remains the gold standard for meeting the first *Gingles* precondition.[159] This data is generally sufficient when it comes to jurisdiction-wide population tasks, such as drawing district boundaries and examining the size and geographic compactness of the minority community, because its sample size is much larger than at the block group level.

As is clear by now, this is not always true for showing racially polarized voting (the second and third *Gingles* preconditions). The limitations on ACS's sample methodology for analysis of small jurisdictions played out dramatically in *Cisneros v. Pasadena Independent School District*, in which the court found that polarization could not be proved due to insufficient data.[160] The court rejected a method similar to BISG—of using surname analysis of actual voters[161]—opting instead to rely solely on ACS VAP data despite its limitations under the circumstances.[162] Doing so led the court to conclude "that there is no evidence of racially polarized voting in the recent endogenous elections."[163] In other words, the ACS CVAP data was not precise enough to demonstrate racially polarized voting in this particular jurisdiction, but alternative methods offering more precision may have been able to do so. The next Part discusses BISG and additional alternative methods that should be considered by Plaintiffs seeking to meet this evidentiary burden under similar circumstances.

---

157. *See* Bernard Grofman & Matt Barreto, *A Reply to Zax's (2002) Critique of Grofman and Migalski (1988): Double Equation Approaches to Ecological Inferences*, 37 SOC. METHODS & RSCH. 599, 608 (2009).

158. *See* Hood, Morrison, & Bryan *supra* note 26, at 547.

159. CHAPA, HENDERSON, NOAH, SCHINKIV, & KENGLE, *supra* note 119, at 2.

160. Cisneros v. Pasadena Indep. Sch. Dist., No. 4:12-cv-2579, 2014 WL 1668500, at *6, *23–24 (S.D. Tex. Apr. 25, 2014).

161. *See infra* Part IV.

162. *Cisneros*, WL 1668500, at *6–8, *11–12.

163. *Id.*, at *21.

IV.
BISG: A PROMISING ADDITIONAL METHOD

As explained in Part III, EI attempts to determine racial voting patterns without knowing (1) the precise vote choice of each voter who cast a ballot or (2) the race or ethnicity for each voter who cast a ballot. The challenge is compounded by the fact that the racial data which is readily available—ACS CVAP data—is of *eligible* voters and not only those who *actually voted* in a given election.[164] If data on actual voters were available, it could be determined how candidate vote choice varies across precincts given the racial demographics of the precinct's voters with greater precision. Indeed, racial vote dilution analyses often rely on data of actual voter turnout.[165] According to political scientist and voting expert M.V. Hood, turnout data on actual voters is always the most preferred source.[166]

As Dr. Hood explains, in some jurisdictions, plaintiffs in voting rights cases have been able to access records for registration and turnout broken down by race and ethnicity and tally the data to show racial polarization.[167] This method has been in use for several decades.[168] One of the first applications of this method was by political scientist Henry Flores, who testified as an expert witness for plaintiffs in *Leal v. San Antonio River Authority*.[169] According to Dr. Flores, the census data was not precise enough because voter turnout for Latinos was far lower than for whites in San Antonio, Texas. Instead, Flores did a manual tally of Spanish surnames on the voter sign-in sheets, by precinct, and correlated that with the number of votes given to Spanish surname candidates. While the census population data did little to reveal voting patterns, Flores' more precise method of looking at the *actual voter list* showed a very strong pattern of racially polarized voting.

Nevertheless, some courts prefer the use of ACS CVAP data in establishing racial demographics.[170] Courts have expressed the concern that Spanish surnames

164. Other means of mitigating these limitations exist but generally involve a dramatic shift in the data captured. For example, experts propose supplementing the potential errors in EI calculations that arise from strategic candidates and interest groups by employing surveys of voters to "measure racial polarization in policy preference, general political ideology, racial attitudes, or oven preferences over race and other candidate attributes as revealed by choices among randomized, hypothetical candidates." *See* Elmendorf, Quinn, & Abrajano, *supra* note 26, 589–90, 675–76.

165. *See* Hood, Morrison, & Bryan, *supra* note 26, at 547–50.

166. *Id.* at 547 ("Given a choice, the order of preference for data type would be turnout, otherwise registration, otherwise CVAP or VAP.").

167. *See id.*

168. Interview with Dr. Henry Flores (Aug. 1, 2021) (discussing Leal v. San Antonio River Auth., CA-5-85-141, SARA SA-85-2988 (July 10, 1985)).

169. *Id.*

170. *See, e.g.*, Westwego Citizens for Better Gov't v. City of Westwego, 906 F.2d 1042, 1045 n.3 (5th Cir. 1990); Reyes v. City of Farmers Branch, No. 07-CV-900-O, 2008 WL 4791498, at *9 (N.D. Tex. Nov. 4, 2008).

are an imperfect proxy for Hispanic self-identification.[171] Latino voters lacking a Spanish surname will be omitted by the method, and non-Latino voters who acquired Spanish surnames (through marriage, for instance) will be counted by the method as a Latino voter.[172] On the other hand, as the Fifth Circuit articulated in *Citizens for a Better Gretna v. City of Gretna*, "*Gingles* . . . suggests flexibility in the face of sparse data."[173] Three years later, the Fifth Circuit further suggested in *Westwego Citizens for Better Gov't v. City of Westwego* that "other probative evidence," including "registered voter data by race," could be considered when census data proved difficult to obtain.[174] Indeed, because analysts are not trying to identify the race of specific individuals, but rather are aggregating totals to precincts and creating percentages, minor errors often cancel each other out, particularly in jurisdictions with extreme racial segregation. Accordingly, many courts have accepted surname analysis of the voter file in certain circumstances.[175] In 2008, the court in *Reyes v. City of Farmers Branch* was presented with a choice between Census data from 2000 and surname estimates from 2006. Finding that the Census data was "out-dated and therefore less likely to be accurate," the court determined that the Spanish surname data was sufficiently probative of the first *Gingles* factor.[176] In *Fabela v. City of Farmers Branch*, the Court used surname data to corroborate the ACS estimate of CVAP for a small geographic area.[177] Citing *Reyes*, the court in *Cisneros* acknowledged: "There is an important need for flexibility in the face of sparse data for vote dilution claims. If census data were unavailable or unreliable, [surname] data would be an appropriate alternative

171. *See, e.g.*, United States v. Alamosa County, 306 F. Supp. 2d 1016, 1022 (D. Colo. 2004); Rodriguez v. Bexar County, 385 F.3d 853, 866 n.18 (5th Cir. 2004) ("[W]ithout a strict showing of its probativeness, Spanish-surname data are disfavored, and census data based upon self-identification provides the proper basis for analyzing Section 2 vote dilution claims in the future.").

172. *See* Cisneros v. Pasadena Indep. Sch. Dist., No. 12-CV-2579, 2014 WL 1668500, at *12 (S.D. Tex. Apr. 25, 2014).

173. 834 F.2d 496, 502 (5th Cir. 1987).

174. *Westwego Citizens for Better Gov't*, 906 F.2d at 1045 n.3.

175. *See* Kumar v. Frisco Indep. Sch. Dist., 476 F. Supp. 3d 439, 478 (E.D. Tex. 2020) (finding that in these circumstances, surname data was "sufficiently probative such that it may be considered"); United States v. Alamosa County, 306 F. Supp. 2d 1016, 1022 (D. Colo. 2004) ("Both experts relied upon Spanish surname analysis to identify Hispanic voters. Spanish surname identification is an accepted means of identifying which voters are likely Hispanic, but it is an imperfect method, particularly in demographically mixed Alamosa County."); Rodriguez v. Harris County, 964 F. Supp. 2d 686, 714 n.14 (S.D. Tex. 2013) ("SSRV is a proxy for the number of Latino registered voters."), *aff'd sub nom*. Gonzalez v. Harris County, 601 F. App'x 255 (5th Cir. 2015); Benavidez v. City of Irving, 638 F. Supp. 2d 709, 725 (N.D. Tex. 2009) ("There are numerous precincts in Irving in which more than 90% of those receiving ballots did not have a Spanish surname, so this methodology was employed to derive estimates of non-Hispanic voters' candidate preferences."); Overton v. City of Austin, 871 F.2d 529, 539 (5th Cir. 1989) ("[The expert] estimated the number of Mexican–American voters, on the other hand, from the number of Spanish surnames on precinct voter registration lists.").

176. Reyes v. City of Farmers Branch, No. 07-CV-900-O, 2008 WL 4791498, *9 (N.D. Tex. Nov. 4, 2008).

177. Fabela v. City of Farmers Branch, No. 10-CV-1425-D, 2012 WL 3135545, *7–8 (N.D. Tex. Aug. 2, 2012).

source of evidence."[178] Even the Department of Justice has relied on this sort of information in its own VRA litigation.[179]

A. How the BISG Method Works

Bayesian Improved Surname Geocoding (BISG) is a statistical modeling method that can be used to overcome limitations that courts have observed in surname data. Political scientists have used Bayesian predictions of racial categorization (also referred to as BISG in the literature) in analysis of voter files and verified its reliability.[180] This technique is commonly used in social science analysis of voting patterns and in other contexts, including health [181] and government. [182] The federal government has employed BISG to assess racial discrimination in consumer finance and voting rights litigation.[183] Most recently, three federal courts have affirmed the reliability of BISG to analyze racially polarized voting in an EI model (in Eastpointe, Michigan, and East Ramapo, New

178. Cisneros v. Pasadena Indep. Sch. Dist., No. 12-CV-2579, 2014 WL 1668500, *9 (S.D. Tex. Apr. 25, 2014) (citing *Reyes*, 2008 WL 4791498, at *9).

179. *See, e.g.*, United States v. Alamosa County, 306 F. Supp. 2d 1016, 1022 (D. Colo. 2004).

180. *See* Kosuke Imai & Kabir Khanna, *Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Records*, 24 POL. ANALYSIS 263, 264 (2016) ("We show that the proposed method reduces the false positive rate among Black and Latino voters to 6% and 3%, respectively, while maintaining the true positive rate at above 80%.").

181. *See, e.g.*, Kevin Fiscella & Allen M. Fremont, *Use of Geocoding and Surname Analysis to Estimate Race and Ethnicity*, 41 HEALTH SERVS. RSCH. 1482 (2006); Marc N. Elliott, Peter A. Morrison, Allen Fremont, Daniel F. McCaffrey, Philip Pantoja, & Nicole Lurie, *Using the Census Bureau's Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities*, 9 HEALTH SERVS. & OUTCOMES RSCH. METHODOLOGY 69 (2009); Dzifa Adjaye-Gbewonyo, Robert A. Bednarczyk, Robert L. Davis, & Saad B. Omer, *Using the Bayesian Improved Surname Geocoding Method (BISG) to Create a Working Classification of Race and Ethnicity in a Diverse Managed Care Population: A Validation Study*, 49 HEALTH SERVS. RSCH. 268 (2014) (validating the use of the BISG method to classify the race/ethnicity of health plan members in health care studies); Stephen F. Derose, Richard Contreras, Karen J. Coleman, Corinna Koebnick, & Steven J. Jacobsen, *Race and Ethnicity Data Quality and Imputation Using U.S. Census Data in an Integrated Health System: The Kaiser Permanente Southern California Experience*, 70 MED. CARE RSCH. & REV. 330, 330 (2012) (describing results that support efforts to use BISG to "conduct studies of racial and ethnic disparities in large health systems").

182. *See* Marc N. Elliott, Kirsten Becker, Megan K. Beckett, Katrin Hambarsoomian, Philip Pantoja, & Benjamin Karney, *Using Indirect Estimates Based on Name and Census Tract to Improve the Efficiency of Sampling Matched Ethnic Couples from Marriage License Data*, 77 PUB. OP. Q. 375 (2013); *see also* United States v. City of Eastpointe, 378 F. Supp. 3d 589, 599 (E.D. Mich. 2019) ("The Consumer Financial Protection Bureau has used [BISG] to predict the race and ethnicity of mortgage applicants.").

183. *See, e.g.*, *City of Eastpointe*, 378 F. Supp. 3d at 599, 613 ("Because the Court is well equipped to weigh Dr. Handley's application of BISG in relation to other evidence submitted in this matter, and finds that the parties have thoroughly informed it of the benefits and risks of BISG, the Court denies without prejudice Defendants' motion to exclude evidence, opinion, and testimony related to the methodology and data, including Defendants' alternative request to appoint an independent expert to evaluate the government's BISG methodology and data."); Brief for Petitioner in Opposition to Defendants' Motion to Exclude Bayesian Improved Surname Geocoding (BISG) Evidence at 8, 19–24, *City of Eastpointe*, 378 F. Supp. 3d at 589 (No. 17-cv-10079).

York).[184] The method relies on a combination of census surname analysis and census block-level racial demographics to provide an overall probability assessment of the voter's race or ethnicity.[185] Voting rights litigants already use each of these measures independently; census data matched to precincts is widely used for understanding precinct racial demographics,[186] and as reviewed above, surname analysis is regularly used against the voter file to understand race and ethnicity.[187] By using both sources of data, it is possible to gain a more precise understanding of voter demographics—two pieces of evidence, instead of just one, provides far more reliable estimates.

BISG analysis begins by undertaking the surname analysis.[188] This is a technique that is commonly used by health scientists, demographers, and sociologists for examining racial and ethnic patterns in health disparities.[189] That is, surname analysis is not new, experimental, or controversial. Rather, it is a well-established method backed by data from the U.S. Census. With respect to voting analysis, where it is possible to obtain a voter file, political scientists have published surname analysis in peer-reviewed political science journals for decades.[190] Surname analysis in BISG starts by taking each last name in the voter file and checking it against the published directories created by the Census Bureau itself.[191] This list, assembled based on research by demographers at the Census Bureau, has created a racial/ethnic probability for each last name in the United

184. *See* Clerveaux v. E. Ramapo Cent. Sch. Dist., 984 F.3d 213 (2d Cir. 2021); NAACP Spring Valley Branch v. E. Ramapo Cent. Sch. Dist., 462 F. Supp. 3d 368 (S.D.N.Y. 2020); *City of Eastpointe*, 378 F. Supp. 3d 589.

185. *See* Imai & Khanna, *supra* note 180, at 264−65.

186. *See* CHAPA, HENDERSON, NOAH, SCHINKIV, & KENGLE, *supra* note 119, at 2.

187. *See* Bernard Grofman & Jennifer R. Garcia, *Using Spanish Surname to Estimate Hispanic Voting Population in Voting Rights Litigation: A Model of Context Effects Using Bayes' Theorem*, 13 ELECTION L.J.: RULES, POLS. & POL'Y 375, 376–77 (2014); CHAPA, HENDERSON, NOAH, SCHINKIV, & KENGLE, *supra* note 119, at 1–2.

188. Marc N. Elliott, Allen Fremont, Peter A. Morrison, Philip Pantoja, & Nicole Lurie, *A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity*, 43 HEALTH SERVS. RSCH. 1722, 1726–27 (2008).

189. *See* Adjaye-Gbewonyo, Bednarczyk, Davis, & Omer, *supra* note 181, at 268; Derose, Contreras, Coleman, Koebnick, & Jacobsen, *supra* note 181, at 330.

190. *See* Matt A. Barreto, Gary Segura, & Nathan Woods, *The Mobilizing Effect of Majority-Minority Districts on Latino Turnout*, 98 AM. POL. SCI. REV. 65, 69 (2004); Matt A. Barreto, *Latino Immigrants at the Polls: Foreign-Born Voter Turnout in the 2002 Election*, 58 POL. RSCH. Q. 79, 82 (2005); Christian Collet, *Bloc Voting, Polarization and the Panethnic Hypothesis: The Case of Little Saigon*, 67 J. POL. 907, 914–915 (2005); Bernard Fraga, *Redistricting and the Causal Impact of Race on Voter Turnout*, 78 J. POL. 19, 25 (2016).

191. *See* Grofman & Garcia, *supra* note 187, at 376.

States, based on the official census records.[192] When a person fills out the census, they record their last name and their self-reported race and ethnicity. Very basic percentages are calculated: for every single person with the last name Garcia on the 2010 Census, 92% indicated they were Hispanic. For every single person with the last name Yu on the Census, 96% indicated they were Asian.[193] The resulting probability estimate for each name then can be cross-referenced with the voter file. So, for nearly every single last name found on a voter file, a surname database can assign it a probability:[194]

Table 3: Probabilities Assigned in Surname Databases

Surname	**% White**	**% Black**	**% Hispanic**	**% Asian**
Barreto	12.6	0.9	83.1	2.8
Cohen	88.9	6.0	3.3	0.7
Dunn	80.1	14.4	2.3	0.8
Collingwood	90.8	4.9	1.6	0.5
Williams	45.8	47.7	2.5	0.5
Johnson	59.0	34.6	2.4	0.5
Zimmer	95.6	0.3	1.9	0.6
Washington	5.2	87.5	2.5	0.3
Gonzalez	4.0	0.4	95.0	0.4
Yu	1.5	0.1	0.7	96.1

For some names—including the authors' own—the surname database correctly assigns a very high probability of the voter's race or ethnicity. While the list has higher probability assignments for Hispanic and Asian names, there are also very commonly occurring names for white and Black voters. There are, however, some names, such as Williams or Johnson, that are common among *both* white and Black voters. Still, the surname analysis is important because even for these voters, it informs us that there is a very low probability that Williams or Johnson is either Hispanic or Asian. With this information in hand, we can move to the next phase of BISG to learn more about voters' racial estimates.

192. *Decennial Census Surname Files (2010, 2000)*, U.S. CENSUS BUREAU (Dec. 15, 2016), https://www.census.gov/data/developers/data-sets/surnames.html [https://perma.cc/9JLV-7NQJ]. An easy-to-use tool that uses the 2010 Census surname list was compiled by Newsday. *How Common Is Your Last Name?,* NEWSDAY, https://projects.newsday.com/databases/long-island/census-last-names [https://perma.cc/W54D-YQ64] (last visited Oct. 14, 2021); *see also* Grofman & Garcia, *supra* note 187, at 375; David L. Word & R. Colby Perkins Jr., *Building a Spanish Surname List for the 1990's–A New Approach to an Old Problem* 1 (U.S. Bureau of the Census, Working Paper No. 13, 1996), https://www.census.gov/content/dam/Census/library/working-papers/1996/demo/POP-twps0013.pdf [https://perma.cc/9N59-MQ8W]; Imai & Khanna, *supra* note 180, at 263.

193. The tool created by Newsday allows users to search by last name and review the racial composition of people with the same last name. *See How Common Is Your Last Name?*, *supra* note 192.

194. *Id.*

The second step of BISG relies on the address of the voter from the publicly available voter file or sign-in sheet from election day.[195] A registrant's physical address is used to correctly assign them to a congressional or state legislative district, as well as a specific voting precinct.[196] Using a procedure known as geocoding, this address information can be cross-referenced with the data from the decennial census at the block level.[197] The census data contains the self-reported race of residents, aggregated to the census block level.[198] According to 2002 U.S. Census report, Black Americans face the highest rates of residential segregation, followed by Hispanics and Asian Americans.[199] Using census data, including the ACS 2016 five-year data, the Washington Post reported that at the neighborhood level "data show most of our neighbors are the same race."[200] Thus, census blocks provide very useful information to assess the probability of a voter's race or ethnicity.

Using census statistics for the racial and ethnic composition for the block in which a voter resides, the block's racial demographic percentages can be used to refine the initial estimate of voter race by surname alone. By using a smaller level of aggregation (i.e., census block), researchers have more precision in their racial estimates. BISG uses two proxy sources of voter race information—a voter's name and where they live—to generate an estimate of their race. By employing the Who Are You (WRU) package in R[201] to estimate the probability that a voter is of a certain race, more accurate vote choice preferences can be inferred from the combination of surname and geolocation data—as opposed to using just one or the other. In one validation exercise, Imai and Khanna demonstrated that the predicted race of the voter very closely matched the actual, self-reported race of the voter for the state of Florida.[202]

---

195. Brian Amos & Michael P. McDonald, *A Method to Audit the Assignment of Registered Voters to Districts and Precincts*, 28 POL. ANALYSIS 356 (2020) (discussing the assignment of voter addresses to election districts).

196. See *id.* at 357–59.

197. *See id.* at 358–59.

198. *See id.*

199. JOHN ICELAND, DANIEL H. WEINBERG, & ERIKA STEINMETZ, RACIAL AND ETHNIC RESIDENTIAL SEGREGATION IN THE UNITED STATES: 1980-2000, 4 (May 9, 2002), https://www.census.gov/library/working-papers/2002/demo/paa-paper.html [https://perma.cc/UC23-HVBJ]).

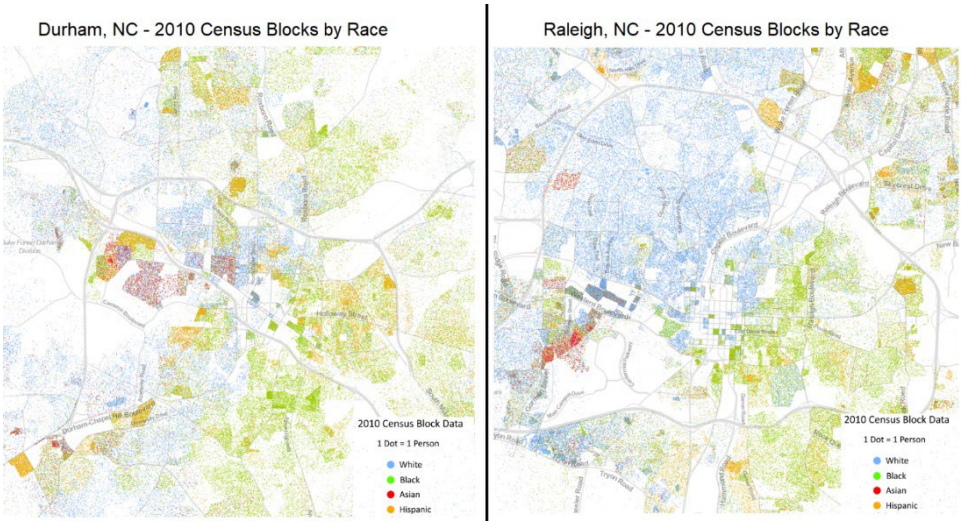
200. Aaron Williams & Armand Emamdjomeh, *America is More Diverse than Ever–but Still Segregated*, WASH. POST (May 10, 2018), https://www.washingtonpost.com/graphics/2018/national/segregation-us-cities/ [https://perma.cc/BT73-ZGHT].

201. R is a computer program and "language and environment for statistical computing and graphics." *What is R?*, R FOUND., https://www.r-project.org/about.html [https://perma.cc/77N9-E785]. It is a free software that any researcher is able to download and utilize. The WRU package that a user can input into R "[p]redicts individual race/ethnicity using surname, geolocation, and other attributes, such as gender and age." Kabir Khanna & Kosuke Imai, Package 'wru' (May 17, 2021) (unpublished manuscript), https://cran.r-project.org/web/packages/wru/wru.pdf [https://perma.cc/XZ8V-T2GW]. The WRU package uses Bayes' Rule to compute the probability of each racial category for any given person.

202. *See* Imai & Khanna, *supra* note 180, at 264.

Some examples will demonstrate how the method works and why it is an improvement over ACS CVAP data alone. For this, we again return to North Carolina to examine racial segregation at the census block level. Figure 3 contains a dot map for Durham and Raleigh for the race and ethnicity within each census block using 2010 decennial census data.[203]

Figure 3: Dot Map of Racial Segregation at Census Block Level in Durham and Raleigh, N.C.



There is clear residential segregation across most of both Durham and Raleigh. Returning to our names above, for a voter with a surname that is scored as 83% likely to be Hispanic, such as Barreto, and who lived in one of the census blocks in Durham that was 82% Hispanic population (e.g., block 2009 in tract 11),[204] BISG would provide an overall score of the racial probabilities that takes both data points into account. The statistical probability of a voter with an 83% Hispanic occurring surname, living in an 80% Hispanic populated census block, being white, Black, or Asian is extremely low. When run through the R package WRU, the two high-probability occurrences reinforce each other to produce an overall Hispanic probability estimate of well over 90%.

203. Dustin A. Cable, *The Racial Dot Map*, UNIV. OF VA., DEMOGRAPHICS RSCH. GRP., https://demographics.coopercenter.org/racial-dot-map/ [https://perma.cc/D7NV-VKWY] (last visited Aug. 26, 2021).

204. *See Hispanic or Latino, and Not Hispanic or Latino by Race*, U.S. CENSUS BUREAU, https://data.census.gov/cedsci/table?q=P2%3A%20HISPANIC%20OR%20LATINO,%20AND%20NOT%20HISPANIC%20OR%20LATINO%20BY%20RACE&g=1000000US370630011002009&tid=DECENNIALPL2020.P2 [https://perma.cc/F2Q6-ZYG3] (last visited Mar. 6, 2022) (linking to underlying data from U.S. Census Bureau table "P2: HISPANIC OR LATINO, AND NOT HISPANIC OR LATINO BY RACE, 2020: DEC Redistricting Data (PL 94-171)").

For surnames that are less unique to one racial group, the census block data greatly helps assign racial probabilities. Take the surname Williams, for which 45.7% of the time is indicated to be white and 47.7% is indicated to be Black, according to Census surname analysis. If we know this voter, Williams, lives in a census block in central Raleigh that is overwhelmingly white—and many of these census blocks are over 90% white—then we have far greater confidence that this Williams is white. However, if a different voter named Williams lived in the eastern/central parts of Durham, which have large Black and Hispanic populations, but almost no white residents—indeed many of these census tracts are less than 5% white—then we would have very high confidence that this Williams is Black. Even though parts of Durham have Black and Hispanic populations living in proximity, surnames such as Williams or Gonzalez do not occur with significant overlap between Blacks and Hispanics. To this point, only 2.5% of people named Williams are Hispanic, and 0.4% of people named Gonzales are Black. Thus, the combination of both surname analysis and census block level data provides a more precise estimate of each voter's race or ethnicity, than just using one method alone.

The method operates much like a familiar probability problem involving a deck of cards. The probability of drawing a red card and a face card will depend on the number of hearts and diamonds (red) in the deck, qualified by the number of jacks, queens, and kings (face cards) in the deck. It uses principles of conditional probability to more precisely estimate racial and ethnic probabilities for a given voter. Since we only have candidate vote choice at the precinct level, we are not concerned with individual level outcomes of BISG, but rather, we want to aggregate the probabilities for each precinct, to correlate with precinct vote totals. Roughly, it is via this method that BISG can infer voter preference—for actual voters—using a combination of surname and geolocation data. We propose that this creates an even more robust picture of voter behavior when implementing EI models.

BISG has at least three significant advantages over relying on ACS's CVAP data in EI models, each directly responsive to the three limitations explained in Part III: the composition of voters included in the data, the survey coverage, and the geographic alignment.[205] First, the population data that the BISG model draws on are *actual voters*, not eligible citizens from which actual voters must be inferred. Recall our example from Durham above. CVAP data includes the entire adult eligible population, in the case of Durham that was 186,727 people,[206] however in the City of Durham mayoral election just 34,867 ballots were cast.[207] Plugging in CVAP race data assumes that racial groups vote in proportion to their size and fails to observe non-linear differences in turnout by race and precinct. Studies have long shown that Black and Latino voters often have lower turnout

205. *Supra* Part III(a).

206. *See Durham County, North Carolina, Sex by Age by Nativity and Citizenship Status*, *supra* note 151.

207. *See* N.C. STATE BD. ELECTIONS, *supra* note 152.

than white voters.[208] Using CVAP data might suggest one precinct is close to 50% Latino, but the actual voter file data might suggest that the same precinct was only 30% Latino among those who cast a ballot on election day. Whenever possible, using data of the actual voters is a more precise way to estimate vote preference.

Second, the ACS 5-year data represents a sample of roughly 10% of households, meaning 90% of households are not interviewed and their race is unknown.[209] BISG, by contrast, uses data from the decennial census, at the census block level, providing data on every household and avoiding associated uncertainty in the estimates.[210] By working with data about everyone in the geographical area, racial estimates are more precise.

Third, ACS CVAP can result in misalignment between voting precinct boundaries and census block group boundaries. Demographers often must trim, collapse, or interpolate how the population in a census block group potentially fits into a voting precinct boundary. We demonstrated this above in Figure 2 by depicting Voting Precinct #19 in Durham County, North Carolina, which includes parts of eight census block groups, only two of which are entirely contained inside the precinct geography, while six census block groups extend well beyond the boundary of Precinct #19.[211] The BISG model, by contrast, only includes people—more specifically, actual voters—who are within each specific precinct, based on the voter file.[212]

This method has proved extremely effective in political science research. Imai and Khana implemented the BISG method to determine if it could improve estimates of race and ethnicity at the precinct level when such data was not available.[213] To validate the success of their model, they compared results of the model to self-reported race, which was available on the voter file in the state of Florida.[214] Using BISG, they first estimated the probability that each voter was white, Black, Latino, or Asian, and then checked the model accuracy by comparing to the self-reported race that roughly 9 million Florida voters indicated on their voter registration forms.[215] They found BISG accurate and capable of improving

208. *See* Paul R. Abramson & William Claggett, *Race-Related Differences in Self-Reported and Validated Turnout*, 46 J. POL. 719, 726 (1984); John R. Arvizu & F. Chris Garcia, *Latino Voting Participation: Explaining and Differentiating Latino Voting Turnout*, 18 HISP. J. BEHAV. SCI. 104, 111 (1996); Kevin Morris, *supra* note 2; *see also* Fraga, *supra* note 2; Daron Shaw, Rodolfo O. de la Garza, & Jongho Lee, *Examining Latino Turnout in 1996: A Three-State, Validated Survey Approach*, 44 AM. J. POL. SCI. 338 (2000).

209. U.S. CENSUS BUREAU, AMERICAN COMMUNITY SURVEY: INFORMATION GUIDE 6 (2017), https://www.census.gov/content/dam/Census/programs-surveys/acs/about/ACS_Information_Guide.pdf [https://perma.cc/S7WD-TYAY].

210. *See* Imai & Khanna, *supra* note 180, at 264–68.

211. *Supra* Part III(b).

212. *Supra* Part IV(a).

213. *See* Imai & Khanna, *supra* note 180, at 264, 270.

214. *Id.* at 267.

215. *Id.* at 267–69.

our understanding of race and ethnicity in voting research.[216] Their work is not alone: "Numerous validation studies have shown that BISG and related methods have an excellent ability to measure race/ethnicity. Concordance between self-reported race/ethnicity and BISG estimates is typically 90 to 96 percent for the four largest racial/ethnic groups—Blacks, Asians/Pacific Islanders, Hispanics, and Whites."[217] One study found that BISG is "19% more efficient than [an earlier method of surname-geocoding] (and 41% and 108% more efficient than single-source surname and address methods, respectively . . .)."[218] Scholars who have worked as experts in voting rights cases have also supported the use of BISG.[219]

B. Aggregate Racial Characteristics, not Individual Level Predictions

BISG works best when researchers look for patterns across racial probabilities as opposed to interpreting the racial classification of a single individual.[220] Some scholars therefore suggest that BISG racial probabilities should be summed and aggregated at the group level.[221] This is not necessarily a limitation in the method because in voting rights lawsuits, analysts are not generally interested in assigning a race to any single individual voter. Instead, they "use the aggregate precinct data to evaluate patterns across precincts, and are therefore more interested in the combined or aggregate racial assignments across precincts."[222] The aggregate data offers a "much more refined read on the racial and ethnic demographics of the voters from one precinct to another because the data is more accurate at an aggregate level."[223]

The aggregation technique smooths out any misclassification that might have occurred at the individual level, and by drawing on the probabilities (instead of a single prediction) it provides a more accurate final count of voters by race and ethnicity. For example, it is possible that an individual voter who self-identifies as Hispanic is only assigned a 30% probability of being Hispanic and a 70% probability of being white. Elsewhere in the same voting precinct, another voter who self-identifies as white is assigned a probability of being 70% Hispanic and 30% white. Whether the reason for the misclassifications is surname or

---

216. *Id.* at 271.

217. Allen Fremont, Joel S. Weissman, Emily Hoch, & Marc N. Elliott, *When Race Ethnicity Data Are Lacking*, 6 RAND HEALTH Q. 1, 2 (2016). *See also* Elliott, Morrison, Fremont, McCaffrey, Pantoja, & Lurie, *supra* note 181, at 69 (describing that when held up against self-reported data, BISG yields an average weighted concordance of 93%).

218. *See* Elliott, Morrison, Fremont, McCaffrey, Pantoja, & Lurie, *supra* note 181, at 69.

219. *See* Hood, Morrison, & Bryan, *supra* note 26, at 547; *see also* Imai & Khanna, *supra* note 180, at 271.

220. *See, e.g.*, Fremont, Weissman, Hoch, & Elliott, *supra* note 217, at 1 ("The BISG method is intended to estimate differences at the group or population level; greater caution should be used in classifying specific individuals' race/ethnicity.").

221. *See id.*

222. Expert Rep. of Matt A. Barreto, Ph.D. & Loren Collingwood, Ph.D. at 3 n.2, NAACP Spring Valley v. E. Ramapo Cent. Sch. Dist., No. 17:17-cv-0894 (S.D.N.Y Jan. 14, 2019).

223. *Id.*

neighborhood demographics, they tend to smooth out when aggregated.[224] That is, the 30% Hispanic assignment for Voter 1 is added to the 70% Hispanic assignment for Voter 2, and BISG reports the precinct to have one Hispanic voter—which is true. In our practice, these misclassifications by BISG are very uncommon, and would not impact the final analysis. When summing the racial probabilities and aggregating to precincts to allow for group level comparison, the BISG model performs extremely well,[225] although empirical research is ongoing in this space.

C. BISG in Court

BISG is a powerful and effective additional tool that addresses the core limitations of using CVAP data, especially in the analysis of smaller jurisdictions. Although its usage in voting rights cases is relatively new, it has quickly been proven reliable, and its reliability has been affirmed by at least two federal courts. When plaintiffs in *United States v. City of Eastpointe* introduced BISG evidence, the court denied the defendants' motion to exclude it on a motion for summary judgment.[226] The court went on to affirm that the plaintiffs had "provided sufficient facts and data to support the reliability of BISG data in this case."[227] Because the case settled, however, the reliability of BISG was never conclusively affirmed.[228]

This matter recently arose again in *NAACP Spring Valley Branch v. East Ramapo Central School District*, in which plaintiffs sought to rely on BISG to show racially polarized voting.[229] In this lawsuit, the two opposing expert witnesses used the same ecological inference (EI) models, but they each used different inputs for the race and ethnicity of voters. The defendant's expert used ACS CVAP data, while the plaintiff's experts relied primarily on BISG analysis

224. When errors do occur, there is no observed pattern in the errors and, essentially, they often cancel each other out. If one person in a voting precinct is mischaracterized as Hispanic, it is often the case that a nearby person who is Hispanic is mischaracterized as not-Hispanic. So, when the data are aggregated to the entire voting precinct, the model still predicts the correct number of Hispanic voters in the precinct. *See* J. Andrew Harris, *What's in a Name? A Method for Extracting Information about Ethnicity from Name*, 23 POL. ANALYSIS 212, 215 (2015).

225. *See* Ioan Voicu, *Using First Name Information to Improve Race and Ethnicity Classification*, 5 STATISTICS AND PUB. POL'Y 1 (2018); CHRISTOPHER T. KENNY, SHIRO KURIWAKI, CORY MCCARTAN, EVAN ROSENMAN, TYLER SIMKO, & KOSUKE IMAI, HARV. UNIV., THE IMPACT OF THE U.S. CENSUS DISCLOSURE AVOIDANCE SYSTEM ON REDISTRICTING AND VOTING RIGHTS ANALYSIS (July 27, 2021), https://arxiv.org/pdf/2105.14197.pdf [https://perma.cc/GY9N-8P4K].

226. 378 F. Supp. 3d 589, 593, 613 (E.D. Mich. 2019).

227. *Id.* at 613.

228. *Justice Department Reaches Agreement with City of Eastpointe, Michigan, Under the Voting Rights Act*, UNITED STATES DEP'T OF JUST. (June 5, 2019), https://www.justice.gov/opa/pr/justice-department-reaches-agreement-city-eastpointe-michigan-under-voting-rights-act [https://perma.cc/FZB8-CS43].

229. 462 F. Supp. 3d 368, 375 n.6, 382 (S.D.N.Y. 2020) (stating that Dr. Barreto, the plaintiff's expert, primarily relied on BISG data).

of the actual voter sign-in data.[230] The court noted that the ability to identify racially polarized voting hinged on whether "BISG is a good data input."[231]

The East Ramapo Central School District is a small jurisdiction that contains only 13 voting precincts in school board elections.[232] Low turnout in these elections means that the ACS CVAP data on all eligible voters, which number 60,000, is an unreliable estimator of the 13,000 to 14,000 people who actually show up to vote.[233] Further, both defense and plaintiffs' experts agreed that Black and Latino voter turnout was lower than whites,[234] which means CVAP likely overestimates how many minorities were voting on election day, introducing bias into the EI estimates.[235]

According to the ACS CVAP data presented by the defense, white voters were cohesive and voted as a bloc, but the data on Black voters was inconsistent and inconclusive.[236] Using BISG on the voter file allowed the plaintiffs to take a more precise look at actual voters and their probability of being Black, revealing a clear pattern of cohesive voting among Black voters.[237] Two immediate trends were clear in the BISG data. First, for both white and Black voters, the confidence interval, or uncertainty estimate surrounding the vote choice prediction, was smaller and tighter, which is evidence of a more accurate prediction. As shown in Table 4 below, for candidate Grossman, CVAP estimated cohesion among white voters, but gave a confidence interval range of 21 points (75 to 96); whereas the BISG estimate for white voters contained a range of just 9 points (69 to 78).[238] Likewise, BISG dramatically improved the reliability of the estimate for Black voters. CVAP produced a confidence range of 79 points (10 to 89) while BISG contained a range of 43 points (5 to 48) for candidate Grossman.[239] Second, BISG detected a clear pattern of cohesive voting among Black voters that CVAP fails to

230. *See id.* at 387.

231. *Id.*

232. *See* E. RAMAPO CENT. SCH. DIST., CORRECTED OFFICIAL BUDGET VOTE RESULTS 2019 (2019), https://www.ercsd.org/site/handlers/filedownload.ashx?moduleinstanceid=47&dataid=19130&FileName=Corrected%20Official%20Budget%20%20Vote%20Results%20-%202019%20-%20Revised%20NTS.pdf [https://perma.cc/HH8K-MGM6]; *E. Ramapo Cent. Sch. Dist.*, 462 F. Supp. 3d at 401 n.45 ("In 2018, the District increased the number of polling places from ten to thirteen.").

233. *E. Ramapo Cent. Sch. Dist.*, 462 F. Supp. 3d at 388 ("Although the District has 60,000 eligible voters, only about 13,000 to 14,000 people actually vote, so using CVAP introduces 'noise . . . influencing who is in a precinct.' . . . Both King's EI and RxC can estimate turnout and incorporate it into the choice percentage, . . . but such estimations do not produce results that are as reliable as the results produced by BISG.").

234. *See id.* at 388–89.

235. *See id.* at 389.

236. Expert Rep. of Dr. John Alford, Ph.D. at 21–23, NAACP Spring Valley v. E. Ramapo Cent. Sch. Dist., No. 17:17-cv-0893 (S.D.N.Y Dec. 21, 2018).

237. *See E. Ramapo Cent. Sch. Dist.*, 462 F. Supp. 3d at 384–87.

238. Expert Rep. of Matt A. Barreto, Ph.D. & Loren Collingwood, Ph.D. at 3 n.2–7, NAACP Spring Valley v. E. Ramapo Cent. Sch. Dist., No. 17:17-cv-0894 (S.D.N.Y Jan. 14, 2019).

239. *Id.*

observe. When running RxC EI, CVAP data estimated a split among Black voters with 53% preferring the white candidate, Grossman, and 47% preferring the Black candidate, Goodwin.[240] The BISG data evidenced that Goodwin was actually preferred by 77% of Black voters, compared to just 23% for Grossman.[241]

Table 4: Comparison of CVAP and BISG Method in Real Election Data in Rockland County, N.Y.

	White		**Black**	
	CVAP	BISG	CVAP	BISG
Grossman (White)	86% (75,96)	74% (69,78)	53% (10,89)	23% (5,48)
Goodwin (Black)	14%% (4,25)	26% (22,31)	47% (11,90)	77% (52,95)
Weissmandl (White)	84% (73,94)	72% (69,75)	56% (11,92)	19% (4,39)
Morales (Hispanic)	16% (6,27)	28% (25,31)	44% (8,89)	81% (61,96)
Lefkowitz (White)	82% (70,93)	67% (62,71)	42% (6,81)	25% (7,46)
Charles-Pierre (Black)	16% (6,29)	31% (27,35)	53% (15,89)	71% (50,92)

After hearing extensive evidence supporting the reliability of BISG, the district court in *East Ramapo* relied upon BISG, calling it "extensively validated by experts"[242] and, "given the unique characteristics of the District . . . a better data set than CVAP for use as an input for ecological inference."[243] The judge dismissed the defendant's criticisms of BISG as "unpersuasive."[244] On appeal, East Ramapo School District argued that the district court abused its discretion in admitting and relying on data derived through BISG; the Second Circuit disagreed.[245] Notably, the Second Circuit upheld the district court's findings that BISG's results could be tested,[246] had been subject to peer review,[247] were reliable,[248] and accepted in the scientific community.[249] Considering the conditions of the case, the Second Circuit found sufficient evidence to support the district court's finding that BISG was the "superior data set."[250] Having secured

240. Expert Rep. of Dr. John Alford, Ph.D. at 18 tbl.3, NAACP Spring Valley v. E. Ramapo Cent. Sch. Dist., No. 17:17-cv-0893 (S.D.N.Y Dec. 21, 2018).

241. *Id.* at 7 tbl.4.

242. *E. Ramapo Cent. Sch. Dist.*, 462 F. Supp. 3d at 383.

243. *Id.* at 387.

244. *Id.* at 389.

245. Clerveaux v. E. Ramapo Cent. Sch. Dist., 984 F.3d 213, 219 (2d Cir. 2020).

246. *Id.* at 234.

247. *Id.*

248. *Id.* at 235.

249. *Id.* at 236.

250. *Id.* at 237.

the Second Circuit's approval, voting rights plaintiffs and local governments should consider BISG safe to rely upon in court.

V.
CONCLUSION

The BISG method shows promise in at least two contexts. First, although not required by *Gingles*,[251] plaintiffs in vote dilution cases should consider using BISG as an additional measure to show racially polarized voting in contexts in which it may yield greater accuracy. One such context is the evaluation of racially polarized voting in smaller jurisdictions. It isn't clear how many vote dilution cases are "just outside of the searchlight of extant litigation," for which ACS survey data does not equip plaintiffs to satisfy the *Gingles* preconditions.[252] By introducing BISG, however, it is plausible that evidence of racially polarized voting in a great deal of cases becomes essentially uncontestable.[253] The enhanced precision of BISG data empowers governments to make informed decisions about its voters and the mandate of Section 2. For these reasons, we propose it is time to move past the theoretical and to the applied, bringing BISG to the fore as a powerful tool for vote dilution plaintiffs and governments alike.

This Article has shown that managing surname data through BISG modeling is an accurate and sometimes necessary innovation in the social science methods that lay the foundation for VRA liability. BISG data is a promising method when evaluating the VRA liability of jurisdictions for which it is difficult to gather the appropriate data using traditional methods. Federal courts have recently deemed BISG admissible and reliable for the first time, laying the groundwork for future reliance on the method. The judicial approval of the method could not come soon enough. BISG has an important role to play as the country enters a redistricting round lacking key protections for minority voters, including Section 5 preclearance, which threatens to lead to new district maps that dilute the minority vote. Especially with the stakes as high as they are, voting rights plaintiffs should be equipped with the most advanced statistical methods. Furthermore, jurisdictions should use BISG when drawing new district boundaries to ensure compliance with the VRA. Armed with this powerful method, voting rights advocates and governments are best equipped to enforce the equal right to vote.

---

251. As stated above, *Gingles* requirements are based on showing vote dilution through methods such as ecological inference using CVAP from the ACS. BISG would be an additional or added optional step one could take and is not required by courts in a Voting Rights Act case. *See* Thornburg v. Gingles, 478 U.S. 30, 49 (1986); *supra* Part III.

252. Justin Levitt, *Citizenship and the Census*, 119 COLUM. L. REV. 1355, 1383 (2019).

253. While not explored in this article, BISG has another potential benefit. Smaller jurisdictions can supplement their VAP or CVAP data and use BISG when drawing new district boundaries to ensure that they are complying with section 2 of the VRA. While VAP and CVAP data are ideal for drawing districts to ensure they comply with the first *Gingles* requirement, a BISG analysis of the actual voters can add confidence that the newly drawn district has enough voters to perform for minority candidates of choice.