

ORIGINAL ARTICLE

The Power of Characters: Evaluating Machine Learning Modified Bayesian Improved Surname Geocoding Inference of Race in Redistricting

John A. Curiel^{*†} and Kevin DeLuca[‡]

[†]Department of Political Science, Ohio Northern University, Ada, OH, USA

[‡]Department of Political Science, Yale University, New Haven, CT, USA

^{*}Corresponding author. Email: j-curiel@onu.edu

Abstract

Identifying racial disparities in policy and politics is a pressing area of research within the United States. Where early work made use of identifying potentially noisy correlations between county or precinct demographics and election outcomes, the advent of Bayesian Improved Surname Geocoding (BISG) vastly improved estimation of race by employing voter lists. Machine Learning (ML) modified BISG in turn offers accuracy gains over the static – and potentially outdated – surname dictionaries present in traditional BISG. However, it is unclear the extent to which ML might substantively alter the policy and political implications of redistricting given its improvements in voter race estimation. Therefore, we ascertain the potential gains of ML modified BISG in improving the estimation of race for the purpose of redistricting majority-minority districts. We evaluate a ML modified BISG program against traditional BISG estimates in correctly estimating the race of voters for creating majority-minority congressional districts within North Carolina and Georgia, and in state assembly districts in Wisconsin. Our results demonstrate that ML modified BISG offers substantive gains over traditional BISG, especially in diverse political geographic units. Further, we find meaningful improvements in accuracy when estimating majority-minority district racial composition. We conclude with recommendations on when and how to use the two methods, in addition how to ensure transparency and confidence in BISG related research.

Keywords: Bayesian Methods, Machine Learning, Redistricting, Racial Representation

1. Redistricting, Race, and Machine Learning

As national level politics increasingly fails to address racial inequities in voter access, representation, and policy outcomes, state governments have become the front lines for political conflict and policy advancement on matters of race. State legislatures are responsible for setting electoral rules in their states, and for creating racial majority-minority districts. Most commonly, the debates over majority-minority districts

center around the plausibility of corrective measures and the degree to which racial disparities can be identified – which is not always possible. State governments vary widely in whether they collect any information on race (Clark, Curiel, and Steelman 2022; Imai and Khanna 2016). Within the context of redistricting, these questions on race can become so severe that state Supreme Courts, such as North Carolina’s, can rule that such racial gerrymandering “impose[s] limits on these legislators’ authority to initiate the process of amending the constitution under these circumstances” given that state legislative maps are “unconstitutionally racially gerrymandered.”¹ Yet such profound decisions on the basis of representation, governance, and ensuing state policies face obstacles in the “race-blind” approach often taken by most data collecting agencies, which severely limits the ability to ascertain whether racial disparities exist and, if so, how to rectify them. States such as North Carolina can only make such decisions due to long-standing legal requirements on the collection of race for voter data as part of currently existing Voting Rights Act (VRA) requirements.²

Therefore, experts in political science, policy, and public health have begun to utilize a diverse set of tools to make up for shortcomings within state data sources where race is not recorded – such as state voter files. Amidst the effort to identify racial disparities, research within the field increasingly makes use of Bayesian Improved Surname and Geocoding (BISG) to impute missing racial data. Although self-reported race data is preferable, BISG methodology provides substantial and significant gains over similar Ecological Inference (EI) methods (Imai and Khanna 2016; King 1997) and is earning wide spread use from the academic community (Decter-Frain *et al.* 2022). Additionally, courts have likewise accepted the estimation of race using surname alone (Barreto *et al.* 2022), with state Supreme Courts – such as Wisconsin’s³ – accepting BISG in vote dilution and civil rights litigation.

Despite the clear improvements in BISG in estimating the first stage of EI issues as related to racial vote preference, there remains room for improvement. Especially in research areas where even slight error can lead to perverse outcomes – such as in redistricting (Hicks *et al.* 2018) – any and all attempts to minimize error are critical. BISG is not a perfect method, with the potential for error especially where the geographic component is heterogeneous, is too aggregated a level, or where a lack of distinct naming patterns exist (Clark, Curiel, and Steelman 2022; DeLuca and Curiel 2022; Voicu 2018). One potential innovation in the application of BISG are developments in machine learning (ML) methods. ML can identify characteristics at the individual character (i.e. letter) level of surname, first, and middle names to improve upon the static surname dictionary employed by traditional BISG estimation (Chaturvedi and Chaturvedi 2020; Decter-Frain 2022; Voicu 2018). Although promising, these ML modifications have not been robustly tested in a manner to overcome two major shortcomings. First, previous validation methods rely upon false negative/positive

1. See *NAACP V. MOORE* 2022-NCSC-99: <https://appellate.nccourts.org/opinions/?c=1&pdf=41699> Additional focused coverage on the case can be found in, Shawna Chen, “N.C. Supreme Court rules against ‘racially gerrymandered’ legislature,” *Axios*, August 19, 2022: <https://www.axios.com/2022/08/19/north-carolina-supreme-court-gerrymander> (accessed October 23, 2022).

2. These requirements stand as of the time of this writing, however, they might be repealed in litigation over Alabama’s Congressional districts’ alleged racial gerrymandering in *Merrill v. Milligan* (2022) 21-1086.

3. *Johnson v. Wisconsin Elections Commission*, 2022 WI 19 – Wis: Supreme Court 2022.

rates that rely upon pluralistic assignment of race – a method that necessarily runs against best practices in ecological inference by exacerbating the specification error via the inflation component (Voss 2004; Palmquist 1993). Second, the application of ML methods have not been applied to a context as stringent as voting rights and vote dilution as related to redistricting, where a strict one-person, one-vote standard exists. Within the context of redistricting and racial gerrymanders, even a minor error can lead to the worst outcome where districts pack a near majority of racial minorities into a racially polarized district such that they are denied descriptive and substantive representation (Grose 2011; Hicks et al. 2018). If ML modified BISG meets redistricting standards, we can have increased confidence in its application broadly, and evidence that validates its use within state and local policy issues.

Therefore, we evaluate the accuracy of ML methods in predicting race relative to traditional BISG. We use the Zest AI Race Predictor (ZRP) machine learning algorithm developed by Zest AI (Matthews et al. 2022), and compare to BISG estimates used at the same geographic level. We evaluate the relative performance within two applications. First, we compare their overall accuracy against the self-reported race data in voter files from Georgia and North Carolina at the congressional district level. We benchmark the performance of ZRP to BISG in the context of redistricting by replicating the analysis in DeLuca and Curiel (2022) using both ZRP and BISG and comparing the results, as stratified by two different race assignment methods – pluralistic and probabilistic. Second, we apply and validate these competing racial estimation methods to state assembly districts within Milwaukee county, Wisconsin. These state assembly districts instigated an intra-party Democratic fight as African-American state assembly members argued that the Democratic map “would make it harder for Black and Latino voters to elect the candidates they want.”⁴ Though Wisconsin lacks data on race within their voter file, we use the next best alternative in the form of aggregated L2 race estimates acquired from Redistricting Data Hub (2022) in our validation.

In our first validation on the congressional redistricting within North Carolina and Georgia, we find that ZRP demonstrates between 6-10% improvements at classifying individuals in most cases relative to traditional BISG. When aggregating to the precinct and district level, ZRP also tends to outperform BISG, though the relative reduction in errors varies significantly by state and across racial groups, given local political geographies. In both states, ZRP performs at least as well as BISG. ZRP shows the most substantive improvements over BISG in Georgia, particularly for estimating the district-level black share of the electorate. In North Carolina, where BISG already performs very well at estimating precinct and district level demographics, ZRP performs just as well or only slightly reduces the already-low error rates. And across states, ZRP tends to reduce error rates the most when using plurality assignment or when estimating the district-level black share of the electorate.

Our second validation in analyzing the competing state assembly maps within Wisconsin demonstrates substantive and decisive gains for ZRP’s ML-modified BISG.

4. Quote from Representative Sylvia Ortiz-Velez (D, Assembly District 8), as quoted in, Bridgit Bowden, “Evers vetoes Republican-drawn redistricting maps,” *Wisconsin Public Radio*, November 18, 2022. <https://www.wpr.org/evers-vetoes-republican-drawn-redistricting-maps> (accessed October 15, 2022).

Where both ZRP and BISG demonstrate stark differences in the degree to which the Wisconsin Republican Party map packs non-white voters into fewer districts, ZRP reduces the error relative to traditional BISG by at least half, and by orders of magnitude in the majority-minority districts of interest. These findings suggest ZRP is far less prone to accidentally “cracking” racial minorities into districts where they lack the ability to elect a member of their preferred race. These results offer an important validation of ML modified BISG as related to redistricting, and more importantly, to the research of race across states and localities within the United States. more broadly. Especially in the study of local governments where traditional EI or even BISG might lead to substantial error (Barreto et al. 2022) or when the pluralistic assignment of race might be absolutely necessary, ML modified BISG can offer substantial gains for those studying state and local politics.

2. BISG: Benefits and Remaining Pitfalls

Originally, Bayesian Improved Surname Geocoding (BISG) was developed in the field of public health in order to address racial disparities in health care (Elliott et al. 2008). In recent years, however, political scientists have adopted BISG in order to estimate racial classifications of individuals and groups in a number of different political contexts, including the context of redistricting (Kenny et al. 2021; DeLuca and Curiel 2022).

BISG uses an individual’s surname and location – typically a Census block, block groups, tract, zip code, or county (Clark, Curiel, and Steelman 2022) – to compute a conditional probability of the individual being of a particular race. Applying Bayes Theorem, one can calculate the conditional probability of individual i , with surname s and geocoded location g , being race r as:

$$P(R_i = r | S_i = s, G_i = g) = \frac{P(G_i = g | R_i = r)P(R_i = r | S_i = s)}{\sum_{i=1}^n P(G_i = g | R_i = r)P(R_i = r | S_i = s)}$$

Typically, increased precision in geography improves race estimation accuracy (Clark, Curiel, and Steelman 2022), and there is variation in how well BISG performs for different racial groups and across states (Imai and Khanna 2016; Dexter-Frain 2022). America’s history of segregation across space and marriage (Massey and Denton 1993) leads BISG to be an especially effective method of studying race within the United States. (Imai and Khanna 2016).

While generally accurate, further reducing errors in BISG would be highly valuable. BISG largely offers gains over traditional EI methods used when studying voter turnout specifically via two components as originally defined by Palmquist (1993) in EI research: reducing the specification and inflation errors. First, BISG employs the especially informative components of name data – as acquired from the Census list on surnames by race (Barreto et al. 2022) – and homogeneous precise geographic units to substantially reduce the specification error. Secondly, even when an observation must be assigned to a single race via pluralistic assignment, BISG inflates the specification error by individual as opposed to all observations within a precinct or county. Where traditional EI research attempts to reduce the error within these two stages to the greatest extent possible (King 1997), these flaws are magnified

in the two stage nature of voting rights EI applications – estimation of turnout and then voter choice. As Barreto et al. (2022) note in identifying the pitfalls in holding up EI research in litigation, these issues become fatally problematic when studying local elections and non-partisan races. Where statewide analyses can approach medium-*N* studies for EI research, sub-state analyses have far fewer voters and precincts to the point that even the most robust EI results are too fragile to meaningfully employ in peer-reviewed research or litigation.

Within the contexts of redistricting, BISG permits estimates of race on individuals within voter lists, which greatly reduces the noise that arises in estimating the racial proportions of those who turnout with only geographic units. Therefore, BISG leads to less uncertainty, and therefore reduces the need to “overly pack” racial minority voters into a few districts as a means to ensure that they can elect a representative of their choice (Lublin 1997; Grose 2011; Hicks et al. 2018). Where employing only census geographic estimates might lead to the desire to create districts where the racial minority population could approach 70 percent (Lublin 1997), BISG can reduce the threshold by at least half (DeLuca and Curiel 2022). While BISG will not solve the long-running normative debate as to whether it is superior to maximize the proportion of minority-influence or majority-minority districts (Steelman and Curiel 2022; Grose 2011; Overby and Cosgrove 1996), it does reduce the uncertainty that in part fuels the divide, thereby leading to potential benefits in substantive and descriptive representation (Pitkin 1967).

Despite clear advances, even the cutting-edge applications of BISG as spearheaded by Imai and Khanna (2016) still see shortcomings. These issues primarily derive from the name component of BISG applications. All BISG packages rely upon surname dictionaries gathered by the Census. Not all surnames can be included in the surname dictionary, with the WRU package containing just under 170,000 surnames. If a name is not included in this list, then a surname merge will fail, leaving a missing imputation or an uninformed prior on the surname (Mateos 2007). Misread characters and list maintenance issues likewise cause merge failures. Second, BISG does not incorporate information on first name, which Voicu (2018) demonstrates substantially improves predictive power. However, even packages modifying BISG, like Bayesian Improved First Name Surname Geocode (BIFSG) in python’s Surgeo package (Weeden, Health, and Naunheim 2022), face the same static and merge issues. Applications of any BISG or BIFSG method with a static dictionary will necessarily be more prone to error as time moves on. These two issues interact to cause general inefficiencies in the data as related to missingness and failed estimations of race. Where BISG cannot estimate race accurately – or at all – it necessarily leads to specification error. Where scholars might likewise apply BISG to local contexts where individuals are assigned to a single race even amidst such uncertainty – pluralistic assignment – BISG applications can start to approach the same type of error that traditionally afflicts normal EI methods (DeLuca and Curiel 2022; King 1997).

We stress that BISG will almost certainly strictly dominate previous EI methods, as discussed at length by Imai and Khanna (2016) and Barreto et al. (2022). However, if the error can be reduced further still and made more adaptable to overcome current structural limitations, then race scholars – especially those within the field of state

and local politics – can attain even greater confidence in their results. These advances would in turn lead to decreased uncertainty, leading to advances in the precision of policy solutions to reduce racial disparities in issues ranging from redistricting and election administration to public health.

ML modified BISG Innovations and Projected Benefits

We seek to gauge the improvements over traditional BISG’s structural issues by employing a machine learning modification to the BISG python package, the Zest AI Race Predictor (ZRP) (Matthews et al. 2022). ZRP addresses the static and missing name issues present in BISG by utilizing the scalable end-to-end tree boosting system – XGBoost – to train on sequences of characters present in first, middle, and last names given the self-reported race, trained upon approximately 30 percent of the data present in voter lists from several states with substantially sized racial minority populations.⁵ ZRP functions by training race predicting models using the joint probability of observing each race conditional on individual characteristics and geographic demographics at the Census Block Group level (Matthews et al. 2022). XGBoost permits the sequential learning from errors and simultaneous models (Chen and Guestrin 2016), with demonstrated robust applications in predicting the race of Twitter users employing surname alone (Wood-Doughty, N. Andrews, and Dredze 2018). Initial checks on the data demonstrate substantial gains in accuracy as applied to correctly identifying correctly identifying African Americans and Hispanic names.⁶

Using ZRP as a modification to BISG therefore overcomes the static nature of the surname dictionary present in BISG by design. The dynamic nature of ZRP likewise means that there will always be some type of informed estimate, which avoids the default uninformed prior for missing data in packages such as WRU.⁷ The simultaneous training on geographic data from the Census and American Community Survey (ACS) additionally means that ZRP could even improve upon the work by Wood-Doughty, N. Andrews, and Dredze (2018) that employs name information alone.

However, it is unclear the extent as of now what these general improvements mean for the study of representation at the state and local level. In order to ensure consistency in future research on race within the field of state and local politics, we need to establish a clear baseline to existing work that makes use of BISG. While it is almost certainly the case that ZRP will offer gains over existing BISG applications, it is important to know to what extent. Additionally, we should ask whether the extent is great enough such that ZRP or similarly designed ML modified BISG should be the standard within the field. An area of explicit focus likewise must consist of how well ML modified BISG improves over traditional BISG in the assignment of voter race. As previously mentioned, work that assigns race to a single category given the most likely estimate is prone to inflating the specification error. Such

5. These states include Florida, Georgia, and North Carolina.

6. Matthews, Kasey. August 4, 2020. “Improving This Algorithm Can Make Lending A Lot Less Racist”, Accessed August 1, 2022.
<https://www.zest.ai/insights/improving-this-algorithm-can-make-lending-a-lot-less-racist>

7. The WRU package uses a naive prior that defaults to a given Census year’s national proportion of the population’s racial categories.

assignment is present in work by scholars such as Enos, Kaufman, and Sands (2019), Grumbach, Sahn, and Staszak (2020), and Lu et al. (2019), even in cases where weighting and aggregating the observations by a given racial category's probability estimate would have been preferable. DeLuca and Curiel (2022) find that within the context of redistricting especially, such pluralistic assignment can lead to substantive and significant error – enough even to accidentally engage in the worst-case scenario “cracking” gerrymander, thereby preventing the representation of racial minorities at the district or legislative chamber level. Benchmarking the gains of ZRP as a case of ML modified BISG against as strict a standard as redistricting would allow us to offer guidance on how and when to use the methods, in addition to enabling a more intuitive comparison to past works.⁸

Therefore, we proceed to first ask what type of gains does ZRP offer over traditional BISG? Second, how do these gains manifest given the two different forms of racial assignment, pluralistic versus probabilistic?

3. First Application: Congressional Redistricting

We validate the ZRP package against the voter files from North Carolina and Georgia as part of a replication and extension of the work by DeLuca and Curiel (2022). In our replication, we employ the ZRP python package against the WRU package created by Imai and Khanna (2016), and extension package zipWRUext by Clark, Curiel, and Steelman (2022) to compare ML-modified estimates of race against traditional BISG estimates. We validate the estimates of each of the methods against the self-reported race within the voter files.⁹ For the purpose of primary figures, we compare the percentage point difference in errors between the WRU BISG estimates and ZRP ML modified BISG.

Our first application employs North Carolina and Georgia for two reasons: feasibility and applicability. The field of BISG validity studies requires a state actually have self-reported race by which to validate against. These almost entirely consist of Voting Rights Act (VRA) covered states with majority-minority district requirements. The issue of monetary costs is also significant: despite federal law requiring accessible statewide reported voter lists, most are not. Most states employ some type of requirement of being a state resident and some type of fee. For example, if we wanted to validate on Mississippi, that would be at least \$1,100, and for Alabama, \$37,000.¹⁰ Therefore, the state of the field necessitates strategic selection of states, even for R1 universities.

Fortunately, North Carolina and Georgia are not only relatively inexpensive, they act as critical cases. Both states effectively started the debate over how to ensure majority-minority districts that actually physically cohere (Monmonier 2001; Curiel and Steelman 2018), a debate that started due to their status as the most populated rural states. Georgia likewise sees a large proportion of rural Black voters, a pattern

8. A prime example of such a work can be found in the article by Steelman and Curiel (2023) in validating different geographic assignment methods' accuracy in relation to geocoded audited voter lists, and project impact in relation to previously published work.

9. We exclude observations with unknown race.

10. Roneka Matheny, “Availability of voter files by state,” *Ballotpedia News*, August 22, 2022. <https://news.ballotpedia.org/2022/08/10/availability-of-voter-files-by-state/> (accessed October 15, 2022).

not present in most methodologically easy applications of BISG, such as those on segregated metropolitan areas (Enos, Kaufman, and Sands 2019). The number of districts required for each state – 13 for North Carolina and 14 for Georgia – make the states exponentially more complex to redistrict relative to smaller states like Alabama or Mississippi (Cho and Liu 2016). Additionally, Lublin (1997) and Grose, Mangum, and Martin (2007) note that the reason why North Carolina and Georgia white Democrats adopted extremely odd-looking and geographically dispersed districts arose from the near impossibility to redistrict additional majority-minority districts that did not sacrifice the reelection prospects of incumbent White Democrats. Therefore, scholars such as Curiel and Steelman (2018), Clark, Curiel, and Steelman (2022), DeLuca and Curiel (2022) make use of these states for earning the highest returns given the cost; if BISG methodologies can survive the stress test by working where its effectiveness should be its weakest, while applied to states where redistricting approaches the limits of ensemble redistricting simulation methods, then it can very likely be applied elsewhere with relative safety. The substantive application to extremely litigious states on majority-minority districts acts as a policy benefit as well, by demonstrating how the methods developed here can address critical real-world redistricting controversies.

In order to perform a one-to-one comparison of the two estimation methods, we first take the privately identifying information (pii) voter file data for North Carolina and run the ZRP command. The ZRP command includes a built-in ZRP geocoder, which places a given individual's address within the following priority look-up sequence: Census Block Group (CBG), Census tract, and finally ZIP code. We therefore ensure that should any differences in estimation arise, it does not arise from different geographic precision.¹¹ Following our estimation of race via ZRP, we run BISG via WRU in R on with the geocoded information from ZRP.¹² Within the North Carolina voter file, we found 83.6% of observations geocoded to the CBG level, and the remaining 16.4% to the ZIP code level. For Georgia, ZRP geocoded 81.3% of observations to CBGs, and the remaining 18.7% to the ZIP code level.

With these estimates, we can then proceed to compare the two posterior allocation methods – plurality versus probabilistic – for aggregating race data up to the precinct, ZIP code, and district level. Using these estimates, we take the difference and first proceed to compare the general error given allocation method, with positive values reflecting greater accuracy on the part of ZRP, and negative values reflecting greater accuracy for BISG as predicted via WRU. At the precinct level, we compare the percentage point difference in error in the predicted number of White and Black voters for the two methods. For ZIP codes, we proceed to calculate the mean error as grouped by the geographic unit against the racial diversity of the area, as calculated with the effective number of races present. The scale for our purposes runs from one to three.¹³

11. DeLuca and Curiel (2022) estimate the race using information at the ZIP code level, following the findings by Clark, Curiel, and Steelman (2022) that using Census block versus ZIP code data produce effectively the same degree of error for White and Black racial estimates.

12. In order to predict race using CBG and ZIP code level data, we made use of the zipWRUext package's "predict_any" command.

13. We reduce the number of races to three – White, Black, and other – to ensure that the results between North Carolina and Georgia are comparable. Additionally, for the purpose of redistricting in North Carolina and Georgia, the majority-minority districts must ensure the representation of Black citizens.

For the district level comparison and potential impact on representation, we then replicate the redistricting simulations by DeLuca and Curiel (2022). Specifically, we use the same redistricting simulation results¹⁴ and merge our estimates of the racial composition of precincts in order to compare BISG and ZRP performance. With these results, we can compare how far each method deviates from the voter file's race data, and with these, calculate the differences in errors.

4. Results: Congressional Redistricting

We first compare the overall performance of WRU estimated BISG against ZRP. More often than not, both the ZRP and BISG imputed races are similar when considering plurality assignment of race (assigning an individual's race to the race with the highest predicted probability). For North Carolina, the ZRP and BISG plurality estimates match 87.57 percent of the time. For Georgia, the plurality match rates are 84.76 percent. Therefore, the ZRP and BISG estimates generally converge on similar estimates when using plurality assignment.

Of the remaining approximately 12.43 percent of observations within North Carolina that are in conflict with each other, the traditional WRU BISG estimates the correct race 23.63 percent of the time. In contrast, the ZRP estimates amidst conflicting estimates produce the correct estimates 62.28 percent of the time. Therefore, amidst conflicting estimates, ZRP is approximately 2.64 times more likely to be correct relative to BISG estimates in North Carolina. Within Georgia, in the 15.24 percent of the cases where BISG and ZRP estimates are in conflict, BISG estimates the correct race 18.04 percent of the time. The Georgia ZRP estimates in turn are correct 61.95 percent of the time amidst such conflict. Therefore, amidst conflicting estimates, ZRP is approximately 3.43 times more likely to be correct relative to BISG estimates in Georgia. These results attest to the power of ZRP's machine learning name component weights over the static surname dictionary seen in the traditional BISG estimates.

Overall, ZRP plurality estimates get 82.9% of respondents correctly classified in the North Carolina voter file, while BISG plurality assignments places 78.0 percent correctly, representing a 6.3 percentage point improvement. In Georgia, ZRP plurality estimates get 76.0 percent of respondents correctly classified while BISG plurality places 69.3 percent correctly classified, representing a 9.7 percentage point improvement.

We next compare the two methods aggregating to precincts. Figure 1 presents the results for each state by race and assignment method. The density plots show the differences in percentage point error rate between BISG and ZRP, such that positive values reflect greater error on the part of traditional BISG relative to the ML modified BISG (ZRP), and negative values indicate the opposite. We see that in North Carolina, weighting and aggregating the probabilities results in the machine learning racial estimation to outperform traditional BISG approximately 56 percent of the time for Whites. In contrast, when assigning race pluralistically, machine learning racial estimation outperforms traditional BISG 84 percent of the time for Whites. Similar results exist for estimating African-American status, with machine learning superior

14. Data and source code are available online (Curiel and DeLuca 2022).

55 percent of the time when aggregating probabilities, and 87 percent for the time when assigned pluralistically.

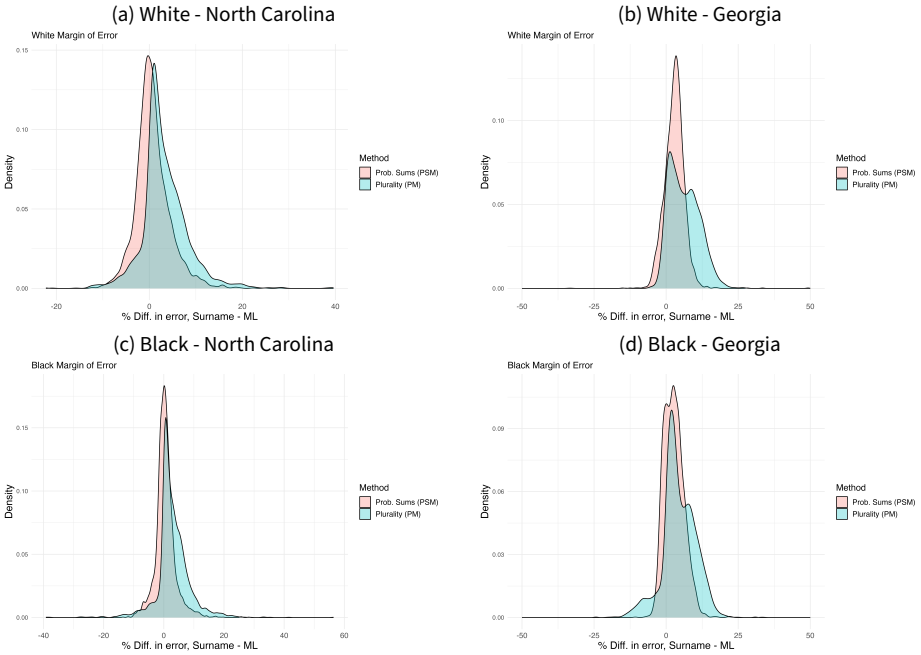


Figure 1. Distribution of Differences Error by Race Assignment and State

Positive values reflect greater error on the part of WRU estimated BISG relative to the ML modified BISG of ZRP, and vice versa for negative values.

We see even greater improvements for ZRP within Georgia: weighting and aggregating the probabilities leads to ZRP outperforming BISG approximately 81.7 percent of the time for Whites, and 90.9 percent of the time when assigning race pluralistically. We get similar results when estimating African-American voters, with ZRP superior 75.4 percent of the time when aggregating probabilities and 86.1 percent for the time when assigned pluralistically.

Overall, ZRP demonstrates improved accuracy over traditional BISG across all precincts. However, the improvements can vary substantially by state. As previously mentioned, the inflation error exacerbates the specification error especially when estimates are less certain. Therefore, we next test whether the degree to which one of the sources of error for both BISG and ZRP – heterogeneous geographic units – leads to less informative priors.

We plot the difference in percentage point error relative to the effective number of races present at the ZIP code level by state. Figure 2 presents the results for North Carolina, while Figure 3 shows the results for Georgia. Both figures categorize the results by assignment method – probability summing or plurality. We see in both

figures that ZRP consistently offers improvements in accuracy relative to traditional BISG as the racial diversity of an area increases. The rate of improvement, however, tends to hit a ceiling around an effective number of races at 1.8 for both North Carolina and Georgia. Around such a point, ZRP tends to have error rates that are about 10 percentage points lower than traditional BISG, and the magnitude of this results is similar across contexts.

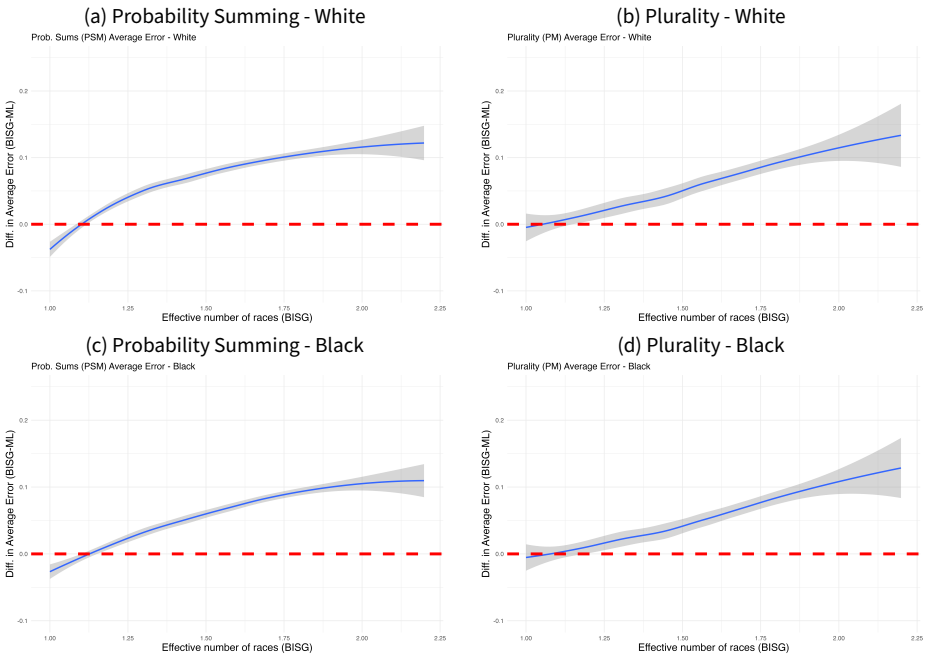


Figure 2. Surname Dictionary BISG v Machine Learning Differences in Error - North Carolina

Y-axis positive values reflect greater error on the part of WRU estimated BISG relative to the ML modified BISG of ZRP, and vice versa for negative values. The x-axis reflects effective number of races, with higher values indicating less informed priors from the geographic units. Shaded area around bars reflect 95 percent confidence interval.

Redistricting Simulations Analysis

We proceed to compare the two methods as aggregated to districts and across 10,000 simulated district maps for each of the two states in Figure 4 and Figure 5. We present the results by state, imputation method, and race. Within an individual plot, we additionally present the error point estimates by assignment method relative to the race as reported within the voter file for a given state. The y-axis reflects the absolute percentage point difference, and the x-axis the proportion of the population of a given race.

Figure 4 presents the results for North Carolina. The results for traditional BISG

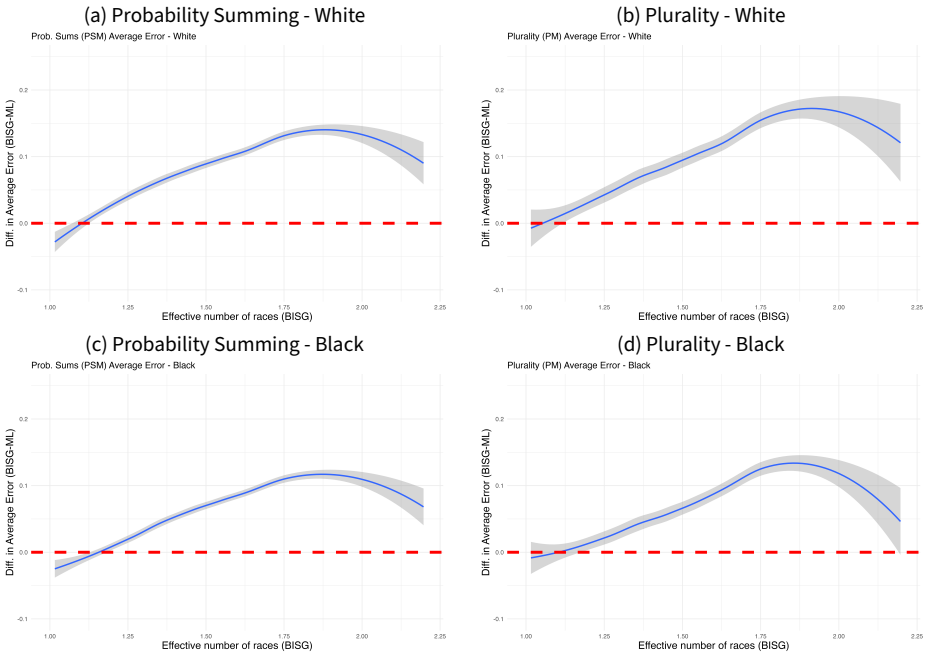


Figure 3. Surname Dictionary BISG v Machine Learning Differences in Error - Georgia

Y-axis positive values reflect greater error on the part of WRU estimated BISG relative to the ML modified BISG of ZRP, and vice versa for negative values. The x-axis reflects effective number of races, with higher values indicating less informed priors from the geographic units. Shaded area around bars reflect 95 percent confidence interval.

estimates in panels (a) and (c) are nearly identical to the original work by DeLuca and Curiel (2022). As originally found, the errors for plurality assignment are significant and substantive in all but the three most heavily concentrated racial minority populated districts. In contrast, the plurality and weighted probability aggregated estimates start to converge towards lack of significant differences for ZRP. The plurality estimates range between approximately two and three percentage points in the estimation of White voters. For the estimation of Black voters, the differences between plurality and weighted probabilities no longer demonstrate significant differences past districts where Black voters comprise over 15 percent of the population. These results suggest that plurality assignment becomes feasible – albeit not necessarily recommended – for redistricting when employing ZRP.

When comparing ZRP to traditional BISG in North Carolina, the plurality ZRP results in panels (b) and (d) show that errors are reduced by about two percentage points. However, the ZRP weighted probability estimates are not significantly different – or in a few cases slightly worse – relative to BISG. Crucially, both methods function similarly where it matters – estimating the composition of the primary majority-minority districts. Both methods converge in reducing the error to under five percentage

points. In summary for North Carolina, ZRP is preferred for plurality assignment, though does not offer improvements over traditional BISG if doing probabilistic assignment.

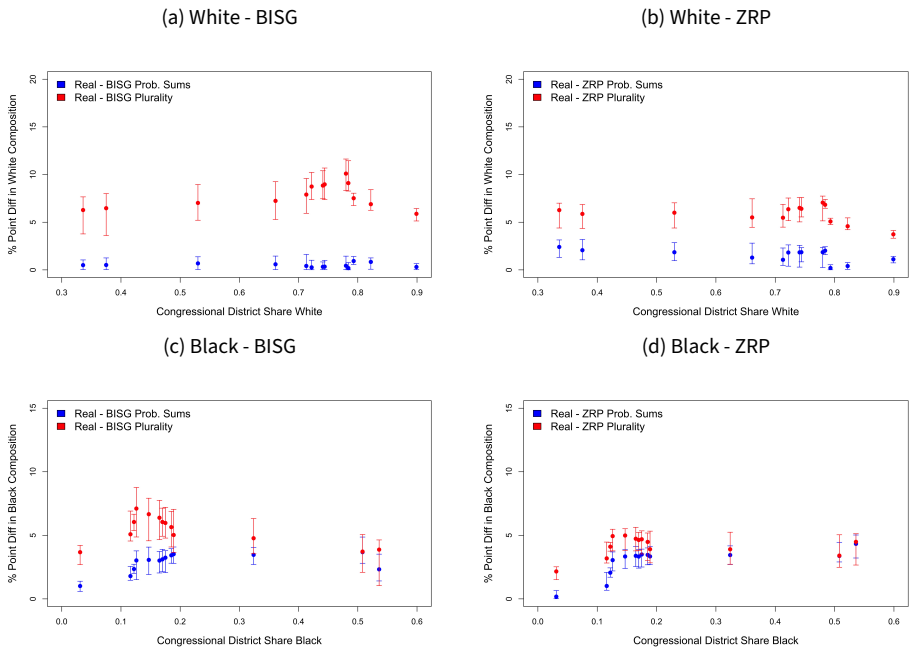


Figure 4. BISG vs ZRP Accuracy - North Carolina

Y-axis reflects the absolute percentage point difference between the voter file race data and racial estimated data. The x-axis the proportion of the population of a given race. Bars reflect 95 percent confidence interval, and the point the median estimate.

Figure 5 presents the results for Georgia. We again see nearly identical estimates in panels (a) and (c) as in the original work by DeLuca and Curiel (2022). Within Georgia, the difference in error between plurality assignment weighted probabilities are significant and substantive in all but two districts for Whites, and the four most heavily concentrated Black districts. Like with North Carolina, we see convergence of plurality and probability weighted estimates for ZRP for estimating Whites within districts (panel b), and no significant differences for estimating Black voters (panel d). ZRP cuts the error rate in half for the purpose of estimating Whites, relative to BISG. While the error rates for Georgia tend to be consistently higher relative to North Carolina when using BISG alone, the estimates become comparable when employing ZRP. For the estimation of Black voters where they comprise under half the district voting population, the ZRP estimates – regardless of assignment method – approach a percentage point difference from the voter file. For the four majority-minority districts with ZRP, the error rates are under 2.5 percentage points. While technically

not significantly different within the majority-minority districts relative to traditional BISG, the error bars are far greater for BISG. Therefore, we see that within Georgia, ZRP produces consistently accurate estimate regardless of assignment method, which is not true for traditional BISG. Therefore, the results suggest that using ZRP in Georgia is preferred over BISG estimates.

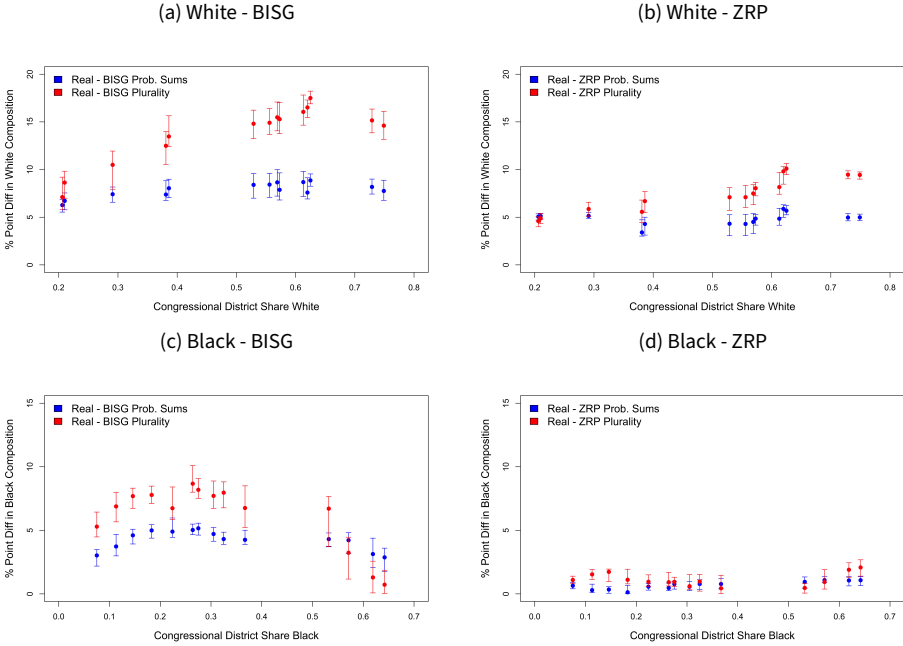


Figure 5. BISG vs ZRP Accuracy - Georgia

Y-axis reflects the absolute percentage point difference between the voter file race data and racial estimated data. The x-axis the proportion of the population of a given race. Bars reflect 95 percent confidence interval, and the point the median estimate.

5. Second Application: Competing State Assembly Maps in Wisconsin

We next proceed to analyze the applicability of ML-modified BISG to traditional BISG within the context of Wisconsin's majority-minority state assembly districts within the Milwaukee metropolitan area. Wisconsin is of interest for a few reasons. First, the Wisconsin Supreme Court in *Johnson v. Wisconsin Elections Commission* (2022) accepted BISG for the purpose of demonstrating racial disparities. Therefore, it provides a useful and immediate baseline for legal adoption in other contexts.

Second, the contention in redistricting within the state after the 2020 Census resulted in a contentious inter- and intra-party conflict. Ultimately, the State Supreme Court chose between two competing maps from which to adopt a minimal-changes

approach for the state assembly districts.¹⁵ However, the lead up to the Democratic proposed map for the state assembly added an additional majority-minority district, which incurred complaints from Milwaukee area assembly members fearing the inability to win reelection.¹⁶

While such concerns are not uncommon, the third reason for interest in Wisconsin is that the state lacks information on race within their voter file. Therefore, unless state assembly members could verify the percentage by race per district with proprietary data – which appears to not have been the case for Democrats who were not directly involved with Governor Evers’s plan – then there would be sufficient uncertainty as to whether the attempt to efficiently spread out racial minority voters might accidentally cause a “cracking” gerrymander. Finally, Wisconsin is of interest as it falls outside the traditionally analyzed Southern VRA-covered states. Therefore, it adds to the geographic diversity in BISG validation studies.

We take advantage of two data sets for the purpose of validating the accuracy of Wisconsin’s registered voter composition. First, while the state of Wisconsin does not publish racial estimates of registered voters, Redistricting Data Hub (2022) purchased and aggregated L2 voter file data for Wisconsin Census Blocks. The L2 data includes internally imputed estimates of race and is considered the standard manner by which to acquire racial estimates in states where voter files lack self-reported race data. A cost prohibitive issue still remained – the Wisconsin voter file itself – which costs \$12,500, one of the most expensive voter files within the nation.¹⁷ However, the MIT Election Data and Science Lab purchased the Wisconsin voter file for November 2020 as part of their Healthy Elections project, and shared with us the data. Therefore, we have a voter file with the necessary information to impute race for both ZRP and BISG, and as credible a “ground truth” as we might achieve in a state that does not collect self-reported race information in their voter file.

We first estimate the WI race data from the WI voter file via the aforementioned ZRP and BISG methods. Upon completion, we aggregate these data – by plurality assignment and weighted probabilities – to the WI electoral ward level into the categories of white and non-white. We proceed to merge these ward aggregated data onto the Wisconsin electoral ward shapefile data, as acquired from the Wisconsin Legislative Technology Services Bureau (2022). We next subset these data to the Milwaukee county municipal wards.¹⁸ We next acquired the three different maps from the internet archived shapefile data from Princeton Gerrymandering Project (2021a) for the Democratic proposed plan, and Princeton Gerrymandering Project (2021b) for the Republican State Assembly proposed plan. We finally spatially overlay

15. Spencer, Douglas. 2022a. “All About Redistricting: Wisconsin State Summary.” <https://redistricting.ills.edu/state/wisconsin/?cycle=2020&level=Congress&startdate=2022-03-03>

16. Bridgit Bowden and Shawn Johnson, “With the help of two Supreme Courts, Republican map prevails,” *Wisconsin Public Radio*, June 1, 2022 <https://www.wpr.org/wpr-reports/mappedout/help-two-supreme-courts-republican-map-prevails> (accessed November 1, 2022).

17. The only other states more expensive are Alabama at \$37,000, California at \$15,000, and Nevada at \$20,000. For a full list of voter file costs by state, see,

Roneka Matheny, “Availability of voter files by state,” *Ballotpedia News*, August 22, 2022. <https://news.ballotpedia.org/2022/08/10/availability-of-voter-files-by-state/> (accessed October 15, 2022).

18. The WI voter file data and ward names had slight variations in names. We hand cleaned these data to confirm a one-to-one match.

both the L2 census block aggregated data and electoral ward ZRP and BISG WI voter file imputed estimates onto the competing maps. Our outcome of interest is the distribution of the difference in the non-white population between the L2 data and ZRP/BISG estimates within the Milwaukee area districts. If we are to see either race imputation method be of use, they should not only follow the rank ordering of districts from least to most white as estimated by L2, but also demonstrate minimal differences.

6. Results for the Wisconsin State Assembly Districts

We present the findings on the differences between the ZRP and BISG estimates in Figure 6 for both the Republican and Democratic proposals. The x-axis plots the districts' racial minority populations, with the 11th district the approximate point of the majority-minority cut point. The L2 data estimates the 11th district from the Republican proposal at a non-white population of 48.3 percent, and the Democratic proposal at 52.5 percent approximately. In the proposed plans, the Republican 10th district contains a non-white population of only 33 percent – likely a situation that falls under a “cracking” gerrymander due to a sizeable yet insufficient racial minority population to select a candidate of their preferred race (Grose 2011). The Democratic plan in turn sees the 10th most district at 49 percent non-white, with the 9th district at 35.5 percent non-white. Across the districts that do fall within the majority-minority category, the Republican proposal ranges from a minimum of 56.5 to a maximum of 73.7 percent non-white. The Democratic plans in turn range from a minimum of 52.5 to 65 percent non-white. These results fall in line with the complaints by Rep. Sylvia Ortiz-Velez that the Democratic map places racial minorities in a precarious position during elections, as the additional majority-minority district notably reduces that margins of registered non-white votes.

We can ascertain the effectiveness of the ZRP v BISG estimates by looking to the y-axis in Figure 6, where positive values reflect an under-estimate by the racial imputation methods, and negative values an over-estimate. Looking to districts one through 10, we see that all imputation methods underestimate the prevalence of racial minorities, with the errors in close proximity for the plurality estimates of ZRP and BISG. The BISG probability weighted estimates place closest to the zero line, hovering around an error of five percentage points, followed by ZRP (i.e. ML-prob.), which places closer to 10 percentage points. However, we see notable differences once we consider the actual majority-minority districts. The BISG plurality and probability weighted estimates range in error from approximately 5 – 10 percentage points in overestimating the percentage of registered non-white voters. These errors increase as the districts become more non-white. The ZRP estimates in turn start to converge with each other for these majority-minority districts in both the Republican and Democratic proposals. The highest error occurs for the 11th district at just over 10 percentage points for the Republican plan, then quickly dropping to an average error of approximately three percentage points for both the Republican and Democratic plans. The most non-white districts see estimates nearly identical to the L2 data.

Notably, we additionally find Rep. Sylvia Ortiz-Velez's state Assembly district 8 –

a ZRP estimated non-white population of 43.5 and 47.7 percent via plurality and probability weighted assignment, respectively. Under traditional BISG, the estimates for the non-white population is approximately 10 percentage points higher. These results in combination with the aforementioned trends suggest that the error between the two methods can be meaningful to an individual state legislator. Both methods place her at the correct rank order of districts, though ZRP/machine learning modified BISG reduces the error and better approximates how marginal the racial minority population is within her district. Therefore, while these results suggest that both methods can be useful absent voter file data on race, it is ZRP that better approximates the gold-standard proprietary data employed by political parties.

7. Conclusion

Overall, ML modified BISG – as estimated with ZRP – weakly dominates traditional BISG. The results for ZRP are at least on par with BISG no matter level of analysis, and is superior in many cases, especially in regards to pluralistic assignment of race. Importantly, we see that ZRP offers the most improvement where the Bayesian prior from a geographic unit is uninformative due to racial diversity. These average gains of 10 percentage points where the effective number of races approaches two suggests that machine learning on names can offer substantive and significant improvements relative to static surname dictionaries present within traditional BISG.

At the congressional district level, we see that state context matters. While the reduction in error rates for North Carolina were minimal relative to BISG, ZRP significantly reduces district-level errors in racial demographic estimates in Georgia. These gains in improvement were across the board, though most crucially in consistently accurate estimates within the majority-minority districts of interest. These results are especially seen within the lesser populated Wisconsin State Assembly districts around Milwaukee. While both ZRP and BISG correctly order the districts, ZRP both reduces the error and does not overestimate the proportion of non-white registrants. Therefore, while both methods improve upon the accuracy and precision of traditional EI methods (Imai and Khanna 2016), ZRP demonstrates greater accuracy than BISG as well. These results are crucial within the context of redistricting, where redistricting bodies frequently engage in cautionary packing (Lublin 1997; Grose 2011) in the event of uncertainty as to racial minority composition and cohesiveness of the electorate. While caution need always be practiced in preventing an accidental cracking racial gerrymander, ZRP can meaningfully reduce uncertainty and hence reduces the risk of unnecessarily packing minority voters into districts.

We therefore recommend the following: first, future research should strive to engage in ML modified BISG estimation of race when possible – especially when small errors can throw off results. While we see that traditional BISG can somewhat hold up within the context of congressional districts' population of several hundreds of thousands of constituents, smaller geographic units – such as state assembly districts – should employ ML modified BISG as a necessity. We see in the context of Wisconsin at least a few districts where substantive differences exist from the actual proportion of minority voters relative to the traditional BISG estimates. Likewise, where research such as that by Curiel and Clark (2021) can still be cited safely in identifying trends

in racial disparities in states like Wisconsin, the impact of error needs be carefully considered in future applications. Within the context of traditional BISG, it might be safer to use a threshold of ten percentage points for congressional redistricting when explicitly designing majority-minority districts, ML-modified methods suggest a five percentage point error rate might not only be used, but even overly cautious. The application to Wisconsin likewise confirms these results, especially when observations are weighted and aggregated by their probabilities. When applied to states where there is reason to believe there might be a difference, users can make use of the effective number of races within a geographic unit as a means to explicitly create a new prior to weight results, should the need arise.

Notably, a limitation of ZRP are the heavier data requirements. ZRP requires at least first and last name, preferably middle name – as part of the ML component – and address information to geocode. These data are not always available, and might be questionably maintained depending on the state’s voter list quality. Additionally, such data are more difficult to request in contexts where pii can be connected with sensitive data, meaning that ZRP might see fewer applications in areas such as public health relative to traditional BISG. However, should such information be available, the ensuing estimates would be consistently more accurate, especially where the geographic units tend to be diverse. We find our results using the ZRP package for python, though the results are consistent with similar ML approaches in estimating race (Wood-Doughty, N. Andrews, and Dredze 2018). Fortunately, ZRP is open source, though presumably other packages that follow a similar method and offer comparable robustness checks against traditional BISG might likewise be used.

Second, should ML modified BISG not be possible, weighting the estimates by probability becomes paramount. As seen in the precinct estimates in Figure 1, whether the error will be greater in ZRP or traditional BISG approached 50–50, in North Carolina at least. This recommendation is in line with previous work (Barreto et al. 2022; Clark, Curiel, and Steelman 2022; DeLuca and Curiel 2022) on BISG, and follows the best practices given the theory of traditional EI methods (Palmquist 1993; King 1997; Voss 2004). While the comparison here between North Carolina and Georgia suggests that the estimates can vary by levels of state intercept shifts, weighting by probabilities offers the most cautious approach. In the event that ML modified BISG is not possible and pluralistic assignment must be used, research ought to report the effective number of races from the atomic geographic units employed as priors. Such information can then be used by authors and reviewers to better weigh their confidence in the results. We see in Figure 6, for example, that as the geographic units become very concentrated in racial minority populations, the difference in error between probabilistic weighting and pluralistic assignment converge toward zero. Research that employs pluralistic assignment and heterogeneous racial composition should be flagged for robustness checks.

For both ML-modified and traditional BISG to remain powerful tools in the estimation of race, it rests upon the current empirical observations and assumptions of meaningful differences in names and geographic residence. While these are long-running patterns within the context of American history, we do recommend semi-regular robustness checks going into the future. Even should America remain segregated,

name preferences will likely continue to adapt.

Insofar as ML-modified BISG or traditional BISG face limitations, it should be within the context of certainty surrounding estimates. There is still room for error, and as seen within the context of Wisconsin, these errors can be meaningful. It is entirely foreseeable that some statistical redistricting consultant might employ either method and tell a representative – such as Rep. Sylvia Ortiz-Velez – that they are overly cautious and should simply adopt barely majority-minority districts that are at best minority-influence districts. Even within the context of peer-reviewed research, there is a general lack of the type of robustness checks that we recommend when talking about uncertainty of estimates (DeLuca and Curiel 2022; Enos, Kaufman, and Sands 2019; Grumbach and Sahn 2020). Likewise, McCarthy (2022) additionally employed pluralistic allocation of race and in analyzing ballot rejection by race in their state audit of Washington State without any discussion on uncertainty – even within the appendix. One of the primary benefits of Bayesian methodology generally is the ability to explicitly state uncertainty in estimates and articulation of priors. Therefore, a stronger effort must be made to caution against treating ML-modified or traditional BISG as some type of magic black box that solves everything. Yes, both are vast improvements over prior methods, though they are not perfect. Treating packages such as ZRP or WRU in highly sensitive contexts such as redistricting can be disastrous.

We conclude that ML modified BISG warrants a place within state and local politics research on race and racial disparities more broadly. BISG already overcomes the greatest limitations within EI research. ML modified BISG with its ability to identify trends at the individual character level overcomes the remaining limitations that might be present in static surname dictionaries. Given the aforementioned acceptance of traditional EI in state level research and litigation, receptivity of surname dictionaries as an improvement over EI in more precise contexts, and BISG methods for the addition of local level application research, ML modified BISG should easily find a home and be welcomed within the academic, legal, and policy fields.

Data Availability Statement. Replication materials are available on SPPQ Dataverse at [placeholder]

Acknowledgements. Thanks to Zest AI Team for their support and development of the ZRP machine learning algorithm evaluated in this paper. Kasey Matthews provided excellent research assistance and technical support.

Funding Statement. This project was supported with an expedited research grant from the Zest AI Team and Schmidt Futures. The authors have been given complete autonomy in developing our analysis and conclusions, and are responsible for all views and any mistakes.

Conflicts of Interest. The authors declare no conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

- Barreto, Matthew, Michael Cohen, Loren Collingwood, Chad W. Dunn, and Sonni Waknin. 2022. A novel method for showing racially polarized voting: bayesian improved surname geocoding. *New York University Review of Law and Social Change* 46 (1).
- Chaturvedi, Rochana, and Sugat Chaturvedi. 2020. It's all in the name: a character based approach to infer religion. <https://arxiv.org/abs/2010.14479>, *arXiv Working Paper*.
- Chen, Tianqi, and Carlos Guestrin. 2016. Xgboost: a scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, 785–794. KDD '16. San Francisco, California, USA: Association for Computing Machinery. isbn: 9781450342322. <https://doi.org/10.1145/2939672.2939785>. <https://doi.org/10.1145/2939672.2939785>.
- Cho, Wendy K. Tam, and Yan Y. Liu. 2016. Toward a talismanic redistricting tool: a computational method for identifying extreme redistricting plans. *Election Law Journal* 15:351–66.
- Clark, Jesse T., John A. Curiel, and Tyler Steelman. 2022. Minmaxing of bayesian improved surname geocoding and geography level ups in predicting race. *Political Analysis* 30 (3): 456–462.
- Curiel, John A., and Jesse T. Clark. 2021. Disparities in poll closures in the age of covid-19: a case study of wisconsin. *Election Law Journal* 20 (4): 345–63.
- Curiel, John A., and Kevin DeLuca. 2022. Replication data for: validating the applicability of bayesian inference with surname and geocoding to congressional redistricting. Harvard Dataverse, <https://doi.org/10.7910/DVN/LXGDWZ>.
- Curiel, John A., and Tyler S. Steelman. 2018. Redistricting out representation: democratic harms in splitting zip codes. *Election Law Journal* 17 (2): 328–53.
- Decter-Frain, Ari. 2022. Benchmarking bayesian improved surname geocoding against machine learning methods. <https://arxiv.org/abs/2206.14583>, *arXiv Working Paper*.
- Decter-Frain, Ari, Pratik Sachdeva, Loren Collingwood, Juandalyn Burke, Hikari Murayama, Matt Barreto, Scott Henderson, Spencer A. Wood, and Joshua Zingher. 2022. Comparing methods for estimating demographics in racially polarized voting analyses. <https://doi.org/10.31235/osf.io/e854z>, *SocArXiv Working Paper*.
- DeLuca, Kevin, and John A. Curiel. 2022. Validating the applicability of bayesian inference with surname and geocoding to congressional redistricting. *Political Analysis*, 1–7. <https://doi.org/10.1017/pan.2022.14>.
- Elliott, Marc N., Allen Fremont, Peter A. Morrison, Philip Pantoja, and Nicole Lurie. 2008. A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity. *Health Services Research* 43 (5p1). issn: 1475–6773, accessed October 26, 2020. <https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1475-6773.2008.00854.x>.
- Enos, Ryan D., Aaron R. Kaufman, and Melissa L. Sands. 2019. Can violent protest change local policy support? evidence from the aftermath of the 1992 los angeles riot. *American Political Science Review* 113 (4): 1012–1028.
- Grose, Christian R. 2011. *Congress in black and white: race and representation in washington and at home*. Cambridge University Press.
- Grose, Christian R., Maurice Mangum, and Christopher D. Martin. 2007. Race, political empowerment, and constituency service: descriptive representation and the hiring of african american congressional staff. *Polity* 39 (4): 449–78.
- Grumbach, J.M., A. Sahn, and S. Staszak. 2020. Gender, race, and intersectionality in campaign finance. *Political Behavior*.
- Grumbach, Jacob M., and Alexander Sahn. 2020. Race and representation in campaign finance. *American Political Science Review* 114 (1): 206–221.

- Hicks, William D., Carl E. Klarner, Seth C. McKee, and Daniel A. Smith. 2018. Revisiting majority-minority districts and black representation. *Political Research Quarterly* 71 (2): 408–23.
- Imai, Kosuke, and Kabir Khanna. 2016. Improving Ecological Inference by Predicting Individual Ethnicity from Voter Registration Record. *Political Analysis* 24 (2): 263–72.
- Kenny, Christopher T., Shiro Kuriwaki, Cory McCartan, Evan T. R. Rosenman, Tyler Simko, and Kosuke Imai. 2021. The use of differential privacy for census data and its impact on redistricting: the case of the 2020 u.s. census. *Science Advances* 7 (41): eabk3283. <https://doi.org/10.1126/sciadv.abk3283>. eprint: <https://www.science.org/doi/pdf/10.1126/sciadv.abk3283>. <https://www.science.org/doi/abs/10.1126/sciadv.abk3283>.
- King, Gary. 1997. *A Solution to the Ecological Inference Problem: Reconstructing Individual Behavior from Aggregate Data*. Princeton, NJ: Princeton University Press.
- Lu, Chao, Yi Bu, Jie Wang, Ying Ding, Vetele Torvik, Matthew Schnaars, and Chengzhi Zhang. 2019. Examining scientific writing styles from the perspective of linguistic complexity. *Journal of the Association for Information Science and Technology* 70 (5): 462–475.
- Lublin, David. 1997. *The paradox of representation: racial gerrymandering and minority interests in congress*. Princeton, NJ: Princeton University Press.
- Massey, Douglas S., and Nancy A. Denton. 1993. *American apartheid: segregation and the making of the underclass*. Cambridge, MA: Harvard University Press.
- Mateos, P. 2007. A review of name-based ethnicity classification methods and their potential in population studies. *Population, Space and Place* 13 (4): 243–63.
- Matthews, Kasey, Austin Li, Christien Williams, Sean Kamkar, and Jay Budzik. 2022. *Zest race predictor*. <https://github.com/zestai/zrp>.
- McCarthy, Pat. 2022. *Performance audit: evaluating washington’s ballot rejection rates*. Technical report 1029711. Office of the Washington State Auditor, February.
- Monmonier, Mark. 2001. *Bushmanders and bullwinkles: how politicians manipulate electronic maps and census data to win elections*. Chicago: University of Chicago Press.
- Overby, Marvin L., and Kenneth M. Cosgrove. 1996. Unintended consequences?: racial redistricting and the representation of minority interests. *Journal of Politics* 58:540–50.
- Palmquist, Bradley Lowell. 1993. Ecological inference, aggregate data analysis of u.s. elections, and the socialist party of america. PhD diss., University of California, Berkeley.
- Pitkin, Hanna. 1967. *The concept of representation*. Berkeley: University of California Press.
- Princeton Gerrymandering Project. 2021a. ‘people’s commission’ state assembly map proposal. Data acquired from the Internet Archived data of the Princeton Gerrymandering Project: <https://web.archive.org/web/20211127154447/https://gerrymander.princeton.edu/>, November.
- . 2021b. *Republican state legislature state assembly map proposal*. Data acquired from the Internet Archived data of the Princeton Gerrymandering Project: <https://web.archive.org/web/2021113031955/https://gerrymander.princeton.edu/>, November.
- Redistricting Data Hub. 2022. *L2 voter file 2020 elections turnout statistics for wi, aggregated to the 2020 census block level*. The Redistricting Data Hub purchased this Voter File from L2, a national Voter File vendor: <https://l2-data.com/>, June.
- Steelman, Tyler S., and John A. Curiel. 2022. Redistricting out descriptive representation: the harmful effect of splitting zip codes on the constituent–representative link. *Journal of Race, Ethnicity, and Politics*, 1–19. <https://doi.org/10.1017/rep.2021.33>.
- . 2023. The accuracy of identifying constituencies with geographic assignment within state legislative districts. *State Politics & Policy Quarterly*, 1–15. <https://doi.org/10.1017/spq.2022.27>.

- Voicu, Ioan. 2018. Using first name information to improve race and ethnicity classification. *Statistics and Public Policy* 5 (1): 1–13.
- Voss, Stephen. 2004. Using ecological inference for contextual research. In *Ecological inference: new methodological strategies*, edited by Gary King, Ori Rosen, and Martin A. Tanner, 69–96. Cambridge, U.K: Cambridge University Press.
- Weeden, Adam, Algorex Health, and Theo Naunheim. 2022. *Surgeo*. <https://github.com/theonaunheim/surgo>.
- Wisconsin Legislative Technology Services Bureau. 2022. *Wi municipal wards*. <https://data-ltsb.opendata.arcgis.com/datasets/LTSB::2012-2020-election-data-with-2020-wards>, July.
- Wood-Doughty, Z., R. Marvin N. Andrews, and M. Dredze. 2018. Predicting twitter user demographics from names alone. *Proceedings of the Second Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media*, 105–11.