

Information Congruence and Redistricting: An Ensemble-Based Test of Whether District Lines Cut Marginalized Communities Off From Local Political Information in the United States, Hungary, and Malaysia

Soham Dwivedi

Independent Researcher

San Jose, California, United States

ORCID ID: 0009-0008-4365-9338

Abstract

Redistricting research quantifies vote dilution but rarely examines how boundaries divide counties and media markets that supply local news. This study asks whether enacted maps fragment marginalized communities' information environments beyond neutral alternatives and whether fragmentation reduces political knowledge. An Information Congruence Index (ICI) measures the share of a small area's county and media-market population assigned to its legislative district. The study draws on fourteen public datasets across three countries. The United States design covers 429 districts across 44 multi-district states, 84,414 census tracts, 210 media markets, and 220,000 ALARM simulated plans, with outcomes that include identifying a representative's party in the 2020–2024 Cooperative Election Studies ($N \approx 180,000$) and House ballot roll-off. The Hungarian design covers 106 single-member constituencies, 3,155 settlements, four elections from 2014 to 2026, and 2022 Roma population shares; the 2024 redrawing of 39 constituencies enables before-and-after comparison. The Malaysian component covers 222 parliamentary seats, 160 administrative districts, and constituency returns with rejected-ballot counts from 2008 to 2022; the 2018 redelineation enables comparison. GerryChain is the reference-plan method for Hungary and Malaysia, and difference-in-differences models are specified that compare areas with ICI declines against unchanged areas, with interactions for race, ethnicity, income, and English proficiency. In the U.S. precinct data, reassigned precincts show 0.86 percentage points more House roll-off than same-incumbent precincts in the same district and election (95% confidence interval 0.41–1.31), and the fitted contrast rises from 0.81 points at a 10% Black population share to 1.14 points at 50%. In Malaysia, constituencies with more than 10% territorial turnover show no distinguishable change in rejected-ballot rates after 2018. Published evidence that majority-minority areas are divided across districts about twice as often, taken together with these estimates, indicates that exposure to the informational costs of redistricting falls disproportionately on marginalized communities.

1 Introduction

A voter who follows local news does not encounter politics through electoral districts alone. Newspapers cover towns and counties; television stations serve audiences that often cross district boundaries; community organizations circulate information through schools, religious institutions, workplaces, and neighborhoods. The geography of these relationships need not coincide with the geography of representation. A resident may repeatedly encounter news about an officeholder who represents nearby communities while receiving little information about the person who represents the resident's own household.

This mismatch provides a reason to study redistricting beyond its effects on the conversion of votes into seats. The placement of a boundary can alter the relevance of information already available in a community. It can also change an outlet's incentive to report on a legislator when that legislator represents only a small portion of the outlet's audience. The resulting cost is not necessarily an absence of political information. It may instead be an abundance of information about national parties and prominent leaders alongside weak knowledge of the local representative whose conduct voters are asked to evaluate.

The central question is whether district boundaries interrupt these information relationships and, if so, whether the interruption falls unevenly across communities. Two mechanisms require separate attention. Reassignment places voters in a district represented by an unfamiliar incumbent, increasing the cost of learning about the local contest. Fragmentation divides a county, neighborhood, or media audience among several districts, potentially reducing the relevance of local reporting over a longer period. Either process may occur without the other. A voter can retain the same incumbent while the surrounding media audience is divided, and a voter can acquire a new incumbent while remaining in a geographically coherent district.

The existing literature supplies substantial evidence for the informational mechanism. Research on newspaper markets links the alignment of readership and congressional districts to political coverage and accountability (Snyder and Strömberg, 2010). Studies of redistricting connect reassignment with candidate knowledge and incomplete voting in congressional contests (Hayes and McKee, 2009, 2012). More recent work extends the congruence argument to state legislatures (Myers, 2026). These contributions make it possible to ask a more precise question than whether redistricting is generally harmful: which relationships are disrupted, for whom, and with what consequences?

This article combines a formal measurement framework with two public-data analyses, a comparative examination of Hungary, and an ensemble-based design for the full test. The U.S. analysis estimates the association between precinct reassignment and House ballot roll-off in the publicly available file linked to Hayes and McKee (2012). The Malaysian analysis joins historical constituency boundaries to election returns and estimates differences in rejected-ballot trends following the 2018 redelineation. The Hungarian discussion examines how changes to constituency boundaries intersect with a media system in which minority information sources face distinct institutional pressures. The ensemble design uses the ALARM Project's simulated congressional plans for the United States and GerryChain reference plans for Hungary and Malaysia to ask whether enacted maps fragment marginalized communities' information environments beyond what neutral alternatives would produce, and it pairs that exposure measure with knowledge and participation outcomes in a difference-in-differences framework. The components address different links in the same argument; their outcomes are therefore interpreted separately.

The principal U.S. result is a positive reassignment coefficient that remains after comparisons are restricted to precincts within the same district and election. The Malaysian estimates do not reveal a comparably clear increase in rejected ballots. This difference is substantively useful. Knowing a representative, completing a congressional ballot, and casting a formally valid ballot are related but distinct behaviors. A theory of informational disadvantage should explain why a boundary change affects a particular outcome, rather than presume that all measures of electoral participation respond in the same way.

2 Literature and theoretical expectations

2.1 The geography of political information

The information available to citizens depends partly on the institutions that produce and distribute it. Jerit et al. (2006) show why political knowledge must be understood in relation to the information environment as well as individual resources. An account based only on education, interest, or income leaves out variation in what people have an opportunity to learn. Electoral geography matters when it changes the relevance of that opportunity.

The economic argument is especially clear for local news. An outlet has stronger incentives to cover a representative when more of its audience lives in that representative's district. Snyder and Strömberg (2010) formalize this relationship using newspaper circulation and congressional boundaries, connecting congruence to coverage, knowledge, and political accountability. Their measure weights the share of an outlet's readers

represented by a legislator by the outlet's importance in the district. It thus captures a relationship between audiences and institutions rather than the visual compactness of a district.

Evidence from state politics strengthens this interpretation. Myers (2026) studies newspaper coverage across the 49 states with partisan legislatures and examines knowledge of state legislators using the 2018 Cooperative Election Study. In that analysis, a one-standard-deviation increase in congruence, approximately 0.19, is associated with a 2.9-percentage-point increase in the probability that a respondent correctly identifies his or her state legislator and a 4.8-point increase in the probability of having heard of the legislator. These are reported results from that study, with its own measure and sample. They provide evidence about the proposed information channel without supplying a coefficient for county–media-market overlap or for representative-party identification in other survey waves.

Television markets show the same pattern. Huber and Tucker (2024) match local television news transcripts to House incumbents and find that a district perfectly congruent with its media market receives about 59 more mentions of its incumbent than a district with minimal overlap; a one-standard-deviation increase in congruence, 0.23, corresponds to about 13.5 additional mentions. Balles et al. (2023) find that smaller television markets, in which a representative's district makes up a larger share of the audience, are associated with more coverage of the representative and greater voter knowledge about him or her. Both studies take the media market as the audience unit, which is the reason the index defined in Section 3 carries a market component alongside the county component.

The local-news literature also suggests that citizens cannot always replace a missing local information source by consuming more national news. Hayes and Lawless (2015, 2018) connect changes in local coverage to knowledge and engagement in House elections. Moskowitz (2021) examines differences in exposure to in-state and out-of-state television markets, linking local political coverage to knowledge and electoral behavior. Newspaper closures are associated with more nationally aligned voting patterns in Darr et al. (2018). Together, these studies support the proposition that the geographical relevance of news matters, even when national political information remains readily available.

2.2 Reassignment, community boundaries, and knowledge

Redistricting can impose a more immediate learning cost by severing an established connection between a voter and an incumbent. In the 1990–92 American National Election Studies panel analysed by Hayes and McKee (2009), 82.4% of respondents drawn into a new incumbent's district could not recall the incumbent's name, compared with 66.4% of respondents who kept their incumbent; the shares unable to recognize the name were 14.2% and 2.9%. Expressed as correct responses, recall is 17.6% among reassigned respondents against 33.6% among those retaining the same incumbent, and recognition is 85.8% against 97.1%. Figure 1 displays these published percentages. The distinction between recall and recognition is important because the latter supplies a cue that reduces the respondent's information burden.

Survey evidence from individual redistricting episodes points the same way. Hood and McKee (2010) find that Georgia voters given a new incumbent by redistricting were less likely to recognize their representative and, when they did not recognize him or her, less likely to vote for the incumbent. Lawrence and Huffmon (2015) report that after South Carolina's 2012 redistricting only about 39 per cent of respondents living in a newly drawn district knew that they had been placed in it. Reassignment therefore removes information that voters had accumulated, and it does so without most of them noticing that anything has changed.

The correspondence between community boundaries and district boundaries can matter even when incumbent continuity is held conceptually separate. Winburn and Wagner (2010) examines county and media-market overlap in relation to political information and participation. Curiel and Steelman (2018) uses the division of ZIP-code areas to examine recognition, contact, and the constituent–representative

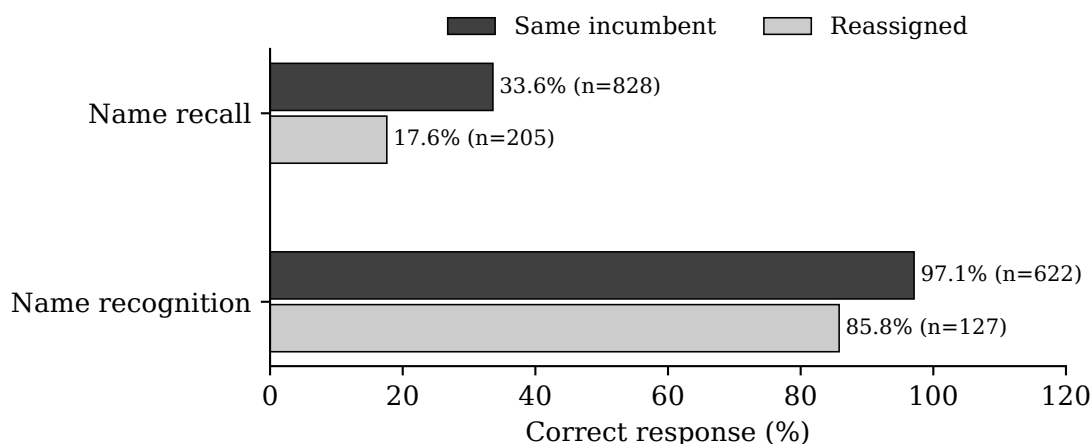


Figure 1. Candidate knowledge and incumbent reassignment. Percentages are redrawn from the descriptive results in Table 1 of Hayes and McKee (2009); sample sizes appear beside the bars. The figure reports published aggregate values, not newly estimated survey coefficients. Recall and recognition use different respondent counts.

relationship. Both approaches treat recognizable communities as settings in which political information circulates. They also suggest that a district's perimeter or area alone cannot identify the communication costs created by its boundaries.

The relevant implication is not that every administrative boundary constitutes a functioning information community. A county can include several distinct audiences, while a newspaper or language community can span multiple counties. Administrative units are useful because they are consistently observed and often organize institutions through which information travels. Their empirical value must nevertheless be established against audience data, outlet coverage, and survey measures of actual exposure.

2.3 Unequal exposure and unequal consequences

An unequal informational burden can arise through two different processes. First, marginalized communities may be more likely to occupy areas that are divided across districts. Second, the same degree of division may have a larger consequence where people have fewer alternative sources of local political information. Residential segregation, income, language access, and dependence on a small number of community institutions could produce either pattern. Separating exposure from susceptibility is therefore essential to identifying the mechanism.

Niven et al. (2022) examine congressional and state House boundaries in 35 states, covering 386 congressional districts and more than 3,600 state House districts. They report that unnecessary overlaps are racially uneven, with substantially more overlap in districts having smaller white population shares. Niven and Solimine (2022) ask the same question at the census-tract level for the congressional maps in force from 2012 to 2021 and find that majority-minority tracts were roughly twice as likely as majority-white tracts to be "castaway," that is, placed in a congressional district other than the one containing most of their county. Both results are evidence of unequal geographical organization across levels of representation. Neither is an ensemble estimate of county or media-market fragmentation, but together they document the exposure side of the distributional concern with the county component of the index defined below.

The consequences of reassignment may also depend on the relationship between voters and representatives. Hayes and McKee (2012) studies the interaction of race, redistricting, and participation, including the role of a coethnic incumbent. That conditional argument cautions against treating the population share of a racial group as a complete explanation of informational disadvantage. A community's organizational resources,

candidate relationships, and access to relevant reporting may change the effect of a boundary disruption.

These considerations yield three expectations. Greater information congruence should be associated with more knowledge of the relevant local representative. Reassignment should increase electoral noncompletion where established incumbent information is lost. The consequences of either process may vary with community resources and minority status, but the direction and magnitude of that variation require separate estimation. A positive average reassignment effect does not, by itself, establish a larger effect for minority voters.

3 Measuring information congruence

3.1 A population-overlap index

Let u denote a small population unit, $d(u, p)$ its legislative district under plan p , $c(u)$ its county or administrative district, and $m(u)$ its media market. Let $N(A)$ be the population residing in geographical set A , measured at a fixed baseline. Define the county and market components as

$$C_{up} = \frac{N\{c(u) \cap d(u, p)\}}{N\{c(u)\}}, \quad M_{up} = \frac{N\{m(u) \cap d(u, p)\}}{N\{m(u)\}}. \quad (1)$$

The Information Congruence Index is

$$I_{up}(w) = wC_{up} + (1 - w)M_{up}, \quad 0 \leq w \leq 1. \quad (2)$$

An equal-weight specification uses $w = 0.5$. The index lies between zero and one when both denominators are positive. It records the share of each surrounding population that is represented in the small area's district. The denominator is the county or market population, not the district population; reversing that denominator would answer a different question about the district's composition.

Small areas that cross a district boundary should be divided at the finest population geography available. If only a fractional crosswalk can be constructed, the area-level value is the population-weighted mean across its constituent fragments. Assigning every resident to the district containing the area's centroid would create avoidable classification error near boundaries, precisely where the exposure of interest is concentrated.

The two components should be reported separately before they are combined. A county component may capture administrative and community ties, while a market component more directly describes an audience. Equal weighting is a transparent starting point, not an empirical claim that the two channels have identical effects. Sensitivity estimates should include county-only, market-only, and alternative-weight specifications. Where circulation or audience shares are observed, an outlet-weighted measure closer to Snyder and Strömberg (2010) provides a useful validation target.

3.2 Interpretation and aggregation

The county component has a simple aggregation property. Suppose county c is divided among districts with population shares $s_{c1}, \dots, s_{cJ}$, where the shares sum to one. A resident in fragment j receives county congruence s_{cj} . The population-weighted county mean is therefore

$$\bar{C}_c = \sum_{j=1}^J s_{cj}^2. \quad (3)$$

This quantity is also the probability that two independently sampled county residents belong to the same legislative district. If the county remains intact, the mean is one. An equal division between two districts produces a mean of one-half; an 80–20 division produces a mean of 0.68. Figure 2 illustrates the arithmetic.

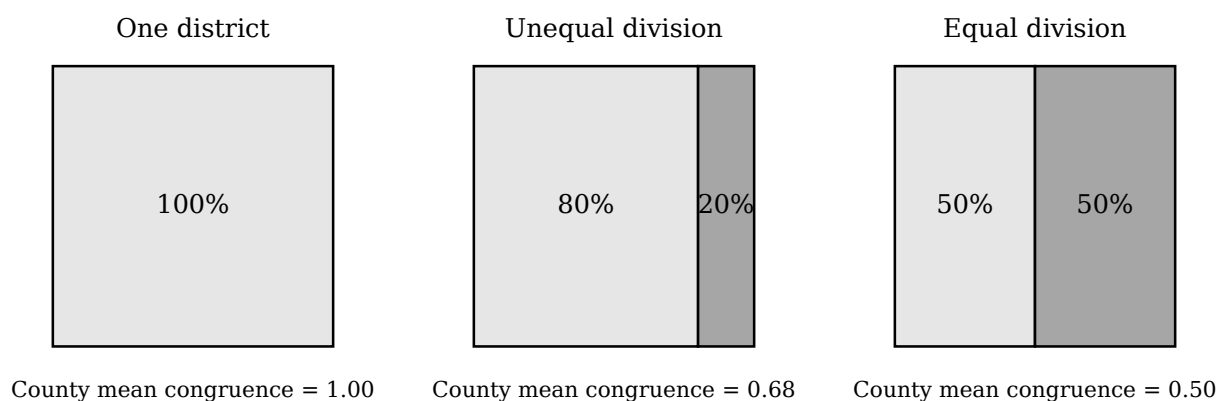


Figure 2. County congruence under three divisions of a population. Rectangles are schematic and assume that area shares equal population shares for illustration. The index itself uses population, not land area. The county mean is the sum of squared district population shares.

This property clarifies both the value and the limitation of the measure. It distinguishes a small peripheral split from an equal division of a community. Yet it does not directly observe the availability, independence, language, or quality of local reporting. A county can remain entirely within one district and have no functioning local newsroom. Congruence is a geographical condition that may facilitate information provision; it is not a direct measure of information consumption.

For population group g , the relevant summary is

$$\bar{I}_{gp} = \frac{\sum_u N_{ug} I_{up}}{\sum_u N_{ug}}, \quad (4)$$

where N_{ug} is the baseline group population in unit u . This weighting describes the average geographical exposure of group members. It does not convert an area-level outcome into an individual-level response. The distinction is central when group composition is related to urban density, residential segregation, or the number of districts that must fit within a large metropolitan area (Robinson, 1950).

3.3 Separating geographical exposure from a causal pathway

Figure 3 summarizes the proposed mechanism. Districting can change both the congruence of an information community and the continuity of its incumbent relationship. Congruence may affect the production and relevance of local coverage, which may influence knowledge. Reassignment may also influence knowledge through familiarity and campaign contact. Population geography and political competition can affect several parts of this sequence, creating confounding that a simple correlation cannot remove.

A convincing test therefore requires more than one outcome. Representative-party identification supplies a relatively low-cost knowledge task, name recognition adds an individual cue, and unaided name recall demands more specific knowledge. House ballot roll-off records a behavioral difference between contests, while a rejected ballot records failure to cast a formally valid vote. These measures should be expected to correlate imperfectly even if they share some informational determinants.

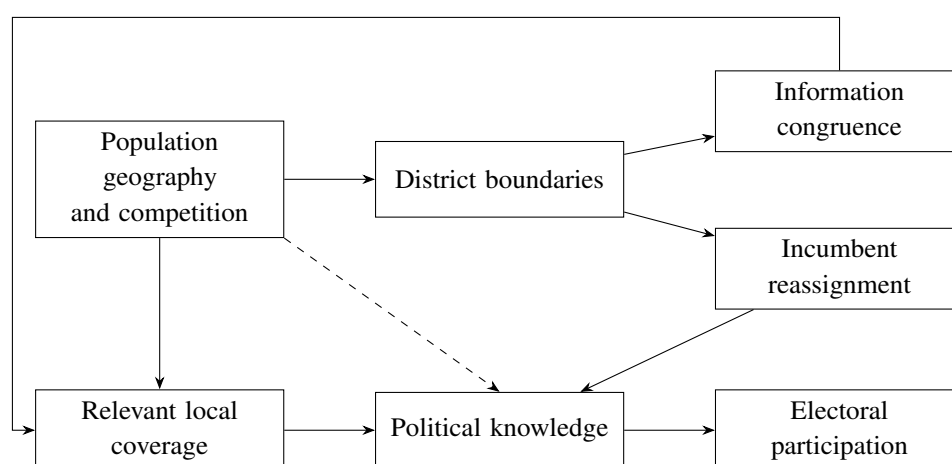


Figure 3. The proposed information mechanism. Arrows express theoretical pathways rather than estimated mediation effects. The dashed path represents additional influences of community characteristics on knowledge. The precinct analysis observes reassignment and roll-off; the Malaysian analysis observes boundary turnover and rejected ballots. Neither alone identifies every link in the diagram.

4 Data and empirical methods

4.1 Data sources

The study draws on fourteen public datasets across the three countries. Table 1 lists them and the part of the study in which each is used. Four enter the estimates reported in Section 5: the precinct file linked to Hayes and McKee (2012), the Malaysian constituency statistics, and the 2003 and 2018 peninsular boundary files. The remainder supply the geographical denominators, reference plans, group populations, and outcome measures of the ensemble and difference-in-differences design specified in Sections 6 and 7. All fourteen are publicly available at the addresses given in the reference list, and the package accompanying the paper records the checksums of the files used for estimation.

4.2 U.S. precinct reassignment and House roll-off

The U.S. reanalysis uses the Stata file publicly linked from Danny Hayes’s research page for the study by Hayes and McKee (2012). The downloaded file contains 76,011 rows. Of these, 40,199 contain zeros in every field and are removed as empty records. The remaining 35,812 rows contain substantive observations. Restricting the analysis to contested elections with a returning incumbent, positive precinct size, and complete values for the outcome, reassignment, composition, district, and election-stratum variables yields 25,750 precinct records. The analytic sample contains 5,547 reassigned records, seven election strata, and 102 district–election cells.

The file available for download differs in coverage from the 65,082 observations described in the published study. The estimates below are consequently a reanalysis of the available file, rather than a reproduction of the article’s full sample or original specification. Its election-stratum codes are retained without assigning geographical labels that are absent from the file’s metadata. Appendix A records the processing decisions and provides the variables needed to reproduce the results.

The outcome is the supplied House roll-off measure, expressed in percentage points. Reassignment is the supplied binary indicator identifying precincts redrawn into a different incumbent’s district. The analysis retains negative roll-off values because a comparison between vote totals in two contests can be negative; such observations are not automatically evidence of invalid records. Uncontested and open-seat races are

Table 1. Public datasets used in the study

	Dataset	Country	Use
1	Precinct file for Hayes and McKee (2012)	United States	Reassignment and House roll-off estimates
2	ElectionData.MY parliamentary statistics (ElectionData.MY, n.d.c; Thevananthan, 2026)	Malaysia	Rejected-ballot estimates
3	ElectionData.MY 2003 peninsular boundaries (ElectionData.MY, n.d.a)	Malaysia	Territorial turnover
4	ElectionData.MY 2018 peninsular boundaries (ElectionData.MY, n.d.b)	Malaysia	Territorial turnover
5	Census 2020 entity tallies and TIGER/Line geography (U.S. Census Bureau, 2022, n.d.b)	United States	Tract, county, and district geography
6	ALARM 50-state redistricting simulations (McCartan et al., 2022; ALARM Project, n.d.)	United States	Reference plans for the ensemble comparison
7	Cooperative Election Study 2020 (Cooperative Election Study, 2021)	United States	Representative-party knowledge
8	Cooperative Election Study 2022 (Cooperative Election Study, 2023)	United States	Representative-party knowledge
9	Cooperative Election Study 2024 (Cooperative Election Study, 2025)	United States	Representative-party knowledge
10	American Community Survey five-year tables (U.S. Census Bureau, n.d.a)	United States	Race, ethnicity, income, and English proficiency by tract
11	DOSM administrative-district population table (Department of Statistics Malaysia, n.d.a)	Malaysia	Administrative denominator; ethnic composition
12	DOSM parliamentary-constituency population table (Department of Statistics Malaysia, n.d.b)	Malaysia	Constituency denominator
13	Census 2022 settlement data, KSH (Hungarian Central Statistical Office, n.d., 2026)	Hungary	Settlement populations and Roma shares
14	National Election Office results, 2014–2026 (National Election Office, n.d.)	Hungary	Constituency assignment and results

Nielsen's documentation of the 210 designated market areas (Nielsen, n.d.) defines the U.S. market geography and is cited as documentation rather than counted as a dataset.

excluded so that the comparison concerns information about a returning incumbent in a contested race.

The principal specification is

$$R_{idt} = \alpha_{dt} + \beta A_{idt} + \gamma' \mathbf{X}_{idt} + \varepsilon_{idt}, \quad (5)$$

where R_{idt} is roll-off, A_{idt} is reassignment, α_{dt} is a district–election fixed effect, and $\mathbf{X}_{idt}$ contains the Black, Hispanic, and other population shares. Composition variables are scaled in ten-percentage-point units. The fixed effects compare precincts exposed to the same district-level contest in the same election, absorbing characteristics that are common within that cell. They do not absorb precinct-specific differences in education, income, campaigning, or political interest.

Two simpler specifications include election-stratum effects, first without and then with composition controls. A fourth specification uses the principal model with the source file's precinct-size variable as a weight. The unweighted model describes the average conditional precinct contrast, whereas the weighted model gives greater influence to larger precincts. Standard errors are clustered by district–election cell, use a finite-sample correction, and form confidence intervals from a t distribution with 101 degrees of freedom.

For heterogeneity, reassignment is interacted separately with the three composition variables. The Black-share interaction measures how the reassignment contrast varies across precinct composition, conditional

on the other terms. It does not estimate a race-specific individual treatment effect. The plotted conditional contrasts hold Hispanic and other shares at their sample means. This choice keeps the graphical comparison interpretable while preserving the distinction between an ecological association and an individual response.

4.3 Malaysian election returns and boundary changes

The Malaysian analysis combines constituency-level election statistics from the Malaysian Election Corpus with digital versions of the 2003 and 2018 peninsular parliamentary boundaries (Thevananthan, 2026; Thevananthan and Chacko, 2025; ElectionData.MY, n.d.c). National descriptive statistics cover the 2008, 2013, 2018, and 2022 general elections, with 222 parliamentary constituencies in each election and 888 constituency–election observations. The panel used for geographical comparisons contains 165 peninsular constituency lineages across four elections, yielding 660 observations. Labuan is excluded from the peninsular design; Sabah and Sarawak are outside the paired geographical files used here.

Ballot totals satisfy the accounting identity

$$\text{ballots issued} = \text{valid votes} + \text{rejected votes} + \text{ballots not returned} \quad (6)$$

for all 888 national records. The outcome is 100 times rejected votes divided by ballots issued. An uncontested constituency in 2008 has no issued ballots and therefore no defined rejection rate; the pre-redelineation comparison uses 164 rather than 165 lineages.

The old and new polygons are transformed to the equal-area coordinate reference system EPSG:6933. All 332 source geometries are valid. Constituency codes are linked across the two maps, with a within-jurisdiction maximum-overlap assignment used to check their correspondence. The overlap check recovers the same one-to-one code matching. This procedure is especially important where constituency names change, since a renamed seat is not necessarily a new geographical unit.

For constituency lineage j , geographical turnover is

$$T_j = 1 - \frac{\text{area}(D_{j,2003} \cap D_{j,2018})}{\text{area}(D_{j,2003} \cup D_{j,2018})}. \quad (7)$$

A value of zero indicates identical polygons, while larger values indicate greater change. This measure concerns territory, not the fraction of voters reassigned and not the Information Congruence Index. A small change in urban land area can involve many residents, whereas a large change in sparsely populated territory may involve few. The area measure permits a reproducible initial comparison but cannot substitute for a population crosswalk.

The principal contrast compares lineages with turnover greater than 10% to those with turnover at or below 10%. Because small differences can reflect digitization as well as substantive redrawing, the comparison group is described as having lower measured turnover, rather than as having unchanged boundaries. The analysis also uses continuous turnover, thresholds of 1%, 5%, and 20%, and weights based on ballots issued in 2013.

For each post-redelineation election t , the first-difference model is

$$Y_{jt} - Y_{j,2013} = \lambda_s + \tau_t \mathbf{1}(T_j > 0.10) + u_{jt}, \quad (8)$$

where λ_s is a jurisdiction fixed effect. A corresponding 2008–2013 model examines the pre-period difference between the same groups. Standard errors are clustered across 13 states and federal territories, with confidence intervals based on t_{12} . The small number of clusters limits the precision of conventional inference. These estimates describe differences in changes under a parallel-trends interpretation; they do not remove changes in constituency composition caused by the redelineation itself.

4.4 Comparison and identification

The two analyses deliberately preserve their distinct outcomes. The U.S. model asks whether reassigned precincts have higher roll-off within a common contest. The Malaysian model asks whether constituencies with greater territorial change experience different trends in ballot rejection. Table 2 specifies the observational units, comparisons, and interpretive scope. A pooled coefficient would have no stable substantive meaning because neither the exposures nor the outcomes are measured on equivalent scales.

Table 2. Empirical components and their inferential scope

Component	Observations and exposure	Outcome or evidence	Principal comparison
United States	25,750 precinct records; incumbent reassignment	House ballot roll-off	Precincts within district–election cells, with composition controls
Malaysia	165 constituency lineages; territorial turnover	Rejected-ballot percentage	Changes from 2013 within states and federal territories
Hungary	106 constituencies and 3,155 settlements; 39 constituencies redrawn in 2024	Boundary reform, minority media conditions, and a before-and-after settlement design	Comparative mechanism evidence and design specification (Section 6)
Congruence framework	Population intersections under enacted and reference plans	County and market overlap	Within-state or within-country ensemble benchmarks

The final two rows specify the comparative design and the measurement and reference-plan framework developed in Sections 6 and 7; they are not additional estimated samples.

Neither fixed effects nor a before-and-after comparison automatically creates a causal design. Precinct reassignment may be related to neighborhood characteristics that also predict participation. Malaysian constituencies may experience different economic or electoral changes, and the composition of their populations may change when the boundary changes. The purpose of the statistical adjustments is to state the comparison clearly and assess its stability, while reserving stronger causal claims for designs that can justify the relevant assignment and trend assumptions.

5 Results

5.1 Reassignment and House ballot roll-off

Reassigned precincts have higher mean roll-off in each of the seven election strata in the public file (Figure 4). The magnitude of the raw difference varies substantially. In the third stratum, reassigned precincts average 9.13% roll-off, compared with 1.66% among same-incumbent precincts. In the second stratum, the corresponding means are 4.21% and 3.40%. The variation makes adjustment for the electoral setting consequential; it would be inappropriate to interpret the pooled descriptive difference as a common effect across elections.

Table 3 reports four specifications using the same 25,750 observations. With election-stratum effects, reassignment is associated with an additional 1.88 percentage points of roll-off. Adding population composition changes the estimate only slightly, to 1.94 points. Introducing district–election effects reduces the coefficient to 0.86 points, with a 95% confidence interval of 0.41–1.31. The reduction indicates that differences across district contests account for a substantial part of the simpler association, although a positive within-contest relationship remains.

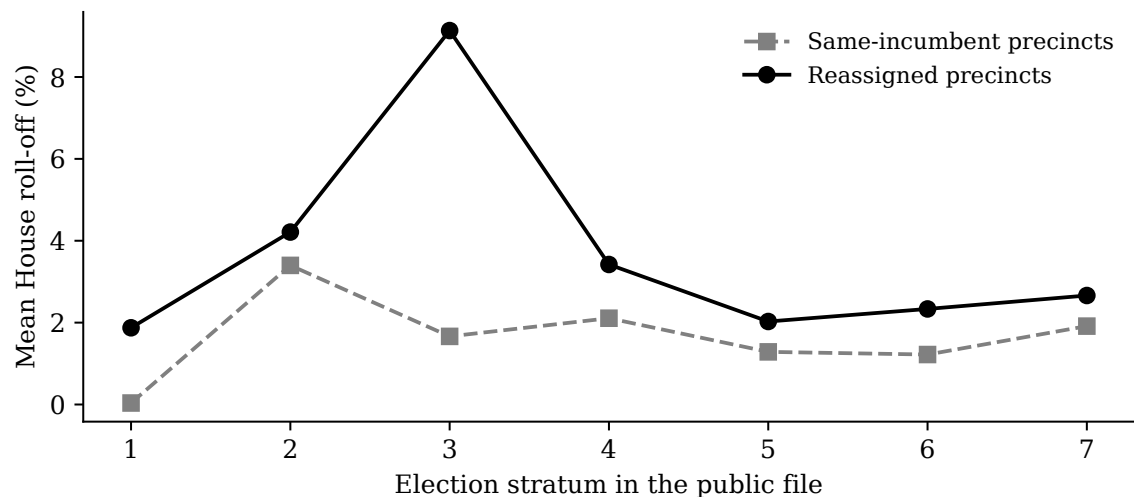


Figure 4. Mean House ballot roll-off by reassignment and election stratum. Means are unweighted and use the common analytic sample. Strata retain the numerical identifiers in the public file. Connecting lines aid comparison across categories and do not imply a chronological sequence.

Table 3. Reassignment and House ballot roll-off

	(1) Strata	(2) Controls	(3) District–election	(4) Size weighted
Reassignment coefficient	1.884	1.935	0.864	0.758
Clustered standard error	0.389	0.380	0.227	0.204
95% CI, lower	1.113	1.181	0.414	0.354
95% CI, upper	2.655	2.689	1.315	1.162
Composition controls	No	Yes	Yes	Yes
District–election effects	No	No	Yes	Yes
Observations	25,750	25,750	25,750	25,750
Clusters	102	102	102	102

Coefficients are percentage-point differences in the supplied roll-off measure. Columns 1–2 include election-stratum effects. Column 4 weights by the source file’s precinct-size variable. All intervals use district–election clustering and t_{101} critical values.

Weighting the principal model by precinct size produces a coefficient of 0.76 points, with a confidence interval of 0.35–1.16. The positive association therefore does not depend on assigning equal weight to precincts of different sizes. Figure 5 displays the coefficients and their uncertainty. Both the weighted and unweighted specifications are compatible with a meaningful electoral cost of losing an established incumbent relationship.

These estimates do not identify the contribution of media fragmentation within that cost. Reassigned precincts may receive different campaign contact, possess less incumbent familiarity, or differ in other characteristics that are not captured by population composition. The direction of the association is consistent with an information disruption, but the model does not observe coverage, knowledge, and roll-off jointly. Its strongest contribution is a transparent estimate of the within-contest reassignment relationship using a public file.

5.2 Population composition and heterogeneity

The reassignment contrast increases modestly with the Black population share in the interaction model, but the slope is imprecisely estimated. A ten-percentage-point increase in Black share is associated with a

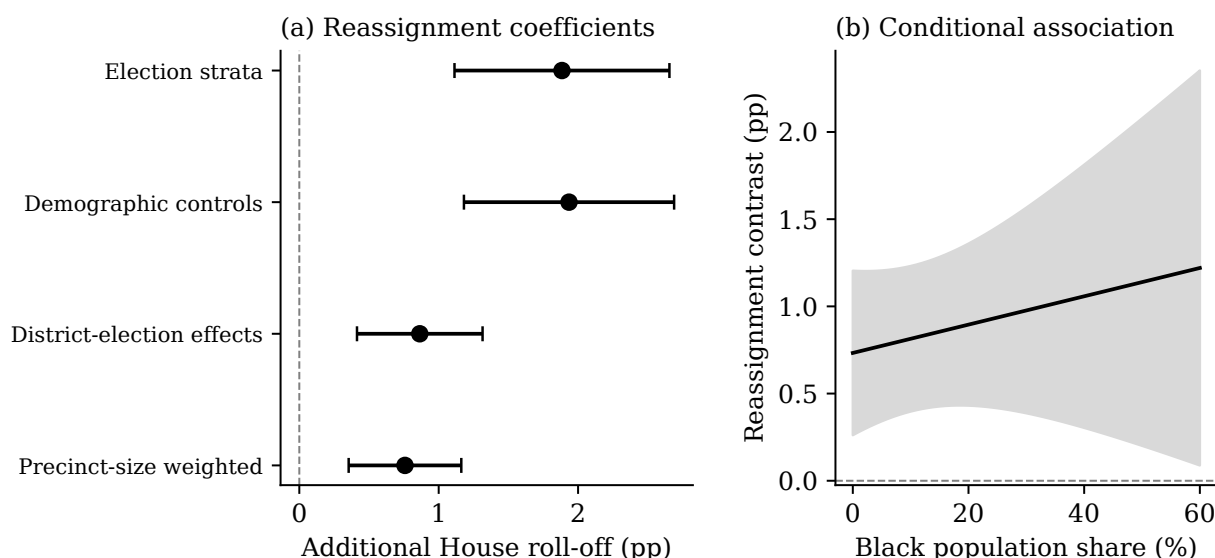


Figure 5. Adjusted reassignment associations. Panel (a) shows coefficients and 95% confidence intervals for the four specifications in Table 3. Panel (b) shows the conditional contrast from a separate composition-interaction model, holding Hispanic and other population shares at their means. The shaded band is a pointwise 95% interval, not a confidence band for the whole curve.

0.081-point increase in the reassignment contrast; the 95% confidence interval ranges from -0.131 to 0.293 , and $p = 0.449$. The data therefore do not distinguish this interaction from zero at conventional levels.

At a Black population share of 10%, the fitted reassignment contrast is 0.81 points, with a confidence interval of 0.40–1.23. At a share of 50%, it is 1.14 points, with an interval of 0.20–2.08. The wider interval at the higher share is visible in Figure 5. Comparing the two point estimates without their covariance and uncertainty would overstate the evidence for differential consequences.

The positive average association and the uncertain interaction answer different questions. The first supports a relationship between reassignment and roll-off in this sample. The second leaves unresolved whether that relationship is systematically stronger in precincts with larger Black populations. Published evidence on racial differences in boundary exposure remains relevant, but exposure disparities cannot be substituted for a directly estimated difference in behavioral response.

5.3 The geography of Malaysian redelineation

Measured territorial turnover is concentrated in a subset of peninsular constituencies. The median is 0.056, and 51 of the 165 lineages exceed the 10% threshold. Figure 6 shows the geographical distribution and the cumulative distribution of turnover. The map describes the scale and location of polygon changes; it does not classify the political intent or population consequences of those changes.

Nationally, the rejected-ballot rate falls from 2.169% in 2008 to 1.543% in 2013, 1.329% in 2018, and 1.014% in 2022 (Table 4). These totals include all 222 parliamentary constituencies and should not be confused with the peninsular estimation sample. The decline is also visible in the unweighted means for both turnover groups. Any claim about a redelineation penalty must therefore concern the relative change between groups, rather than an increase in the national level of rejected ballots.

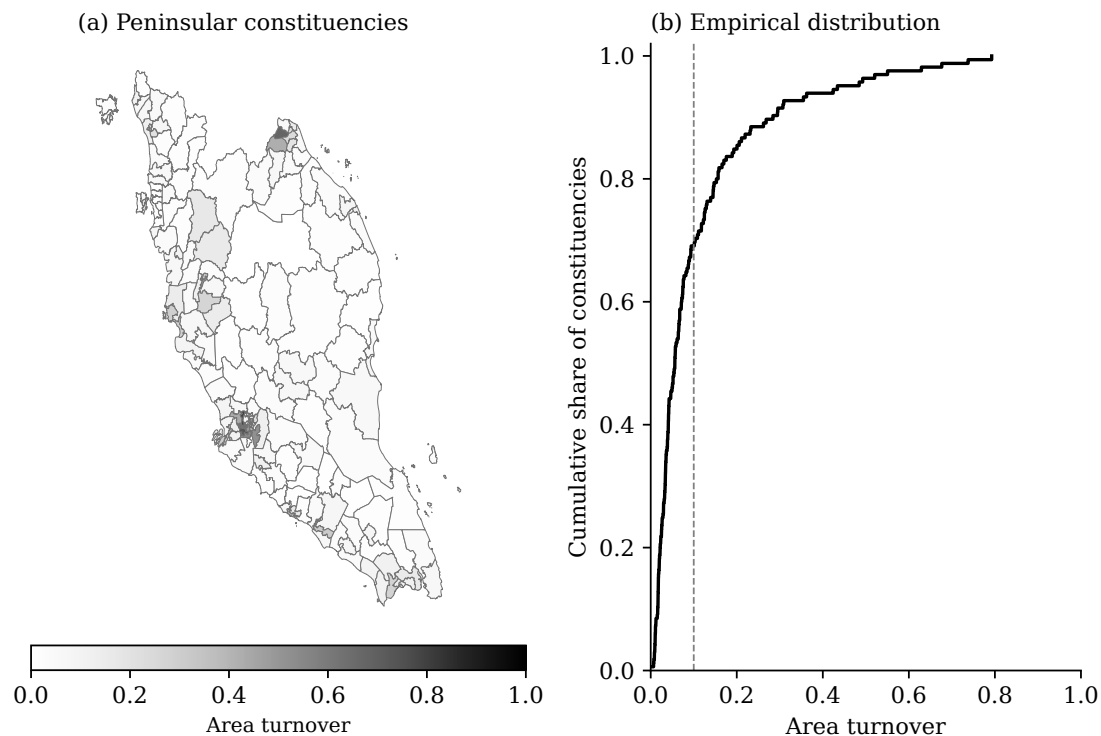


Figure 6. Territorial turnover between the 2003 and 2018 peninsular parliamentary maps. Darker shading indicates larger values of Equation 7. The cumulative distribution contains 165 constituency lineages; the dashed line marks the 10% threshold. Areas are calculated in an equal-area projection. Sources: ElectionData.MY (n.d.a) and ElectionData.MY (n.d.b); author's calculations.

Table 4. National Malaysian parliamentary ballot totals

Election	Constituencies	Ballots issued	Rejected ballots	Rejected (%)
2008	222	8,161,152	176,995	2.169
2013	222	11,256,545	173,661	1.543
2018	222	12,300,067	163,504	1.329
2022	222	15,817,307	160,340	1.014

The rate is the ratio of national rejected ballots to national ballots issued, multiplied by 100; it is not an average of constituency percentages. Source: ElectionData.MY (n.d.c); author's calculations.

5.4 Rejected-ballot trends and sensitivity

The 10% threshold estimate for 2013–2018 is -0.041 percentage points, with a 95% confidence interval of -0.123 to 0.040 . For 2013–2022, it is 0.067 points, with an interval of -0.053 to 0.187 . The pre-period estimate for 2008–2013 is close to zero, although its interval is wide enough to permit nontrivial differences in underlying trends. These results provide no precise evidence that greater territorial turnover increased ballot rejection in the first election under the new boundaries.

The continuous specification gives the same substantive impression. A 0.10 increase in turnover is associated with a 0.001-point difference in the 2013–2018 change and a 0.027-point difference in the 2013–2022 change. Both confidence intervals include zero. Weighting by pre-redelineation ballots also leaves the main estimates small and imprecise. These results make it difficult to attribute the absence of a clear binary-threshold result solely to the choice of the 10% cutoff.

Figure 8 reports the remaining threshold models and a leave-one-jurisdiction-out diagnostic. The 1% threshold produces a positive 2022 estimate of 0.259 points, with a confidence interval of 0.012–0.506.

Table 5. Boundary turnover and changes in Malaysian rejected-ballot rates

Specification	2008–2013	2013–2018	2013–2022
Turnover greater than 10%	0.004	−0.041	0.067
95% confidence interval	[−0.170, 0.177]	[−0.123, 0.040]	[−0.053, 0.187]
Continuous turnover, per 0.10	0.003	0.001	0.027
95% confidence interval	[−0.049, 0.055]	[−0.035, 0.036]	[−0.016, 0.070]
10% threshold, ballot weighted	0.022	−0.018	0.081
95% confidence interval	[−0.138, 0.181]	[−0.113, 0.076]	[−0.050, 0.211]
Constituency lineages	164	165	165
Jurisdiction clusters	13	13	13

All coefficients are percentage-point differences in changes, with jurisdiction fixed effects. Continuous coefficients correspond to an increase of 0.10 in territorial turnover. Weighted models use ballots issued in 2013. Intervals use jurisdiction clustering and t_{12} critical values.

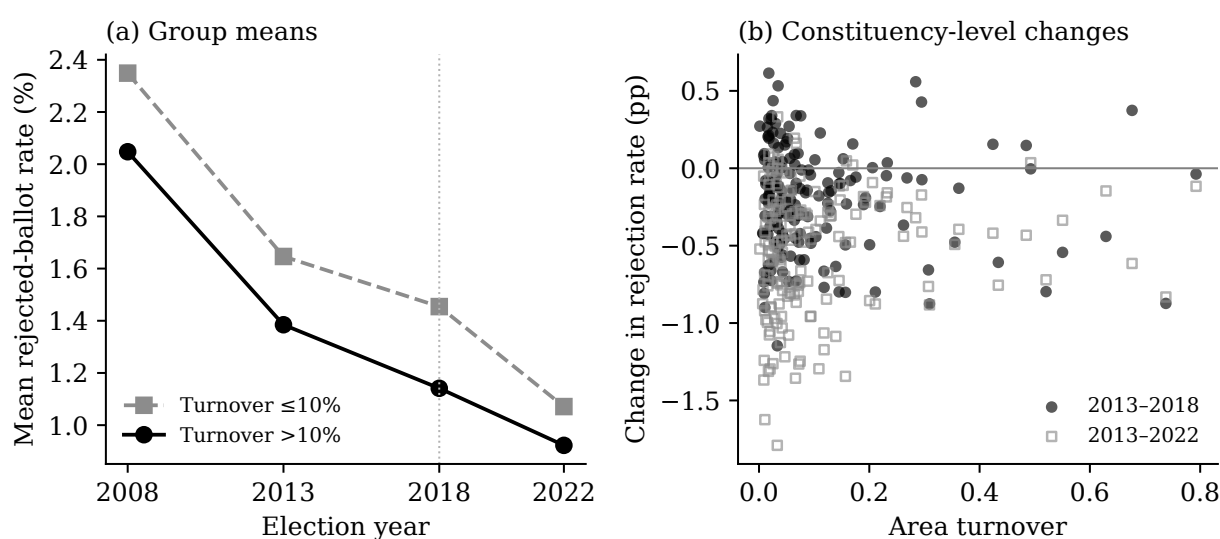


Figure 7. Rejected-ballot trends and territorial turnover. Panel (a) presents unweighted constituency means for higher- and lower-turnover groups; the vertical line marks the 2018 election. Panel (b) plots constituency changes relative to 2013 against continuous turnover. Descriptive group means are not adjusted for jurisdiction.

That comparison leaves only seven lower-turnover lineages and is not reproduced by the 5%, 10%, or 20% thresholds. It should therefore be interpreted as sensitivity to the composition of a small comparison group, rather than selected as the preferred result because it crosses a conventional significance threshold.

Omitting one jurisdiction at a time leaves the 2018 point estimate between -0.077 and -0.030 points and the 2022 estimate between 0.028 and 0.090 . No single jurisdiction reverses the sign of the principal point estimates. The exercise does not establish parallel trends or solve the small-cluster problem, but it shows that the main estimates are not driven entirely by one state's observations.

6 Comparative interpretation: Hungary and Malaysia

6.1 Hungarian boundaries and minority information sources

Hungary provides a relevant setting for examining the conjunction of constituency changes and a politically constrained information environment. The 106 single-member constituencies drawn under Act CCIII of

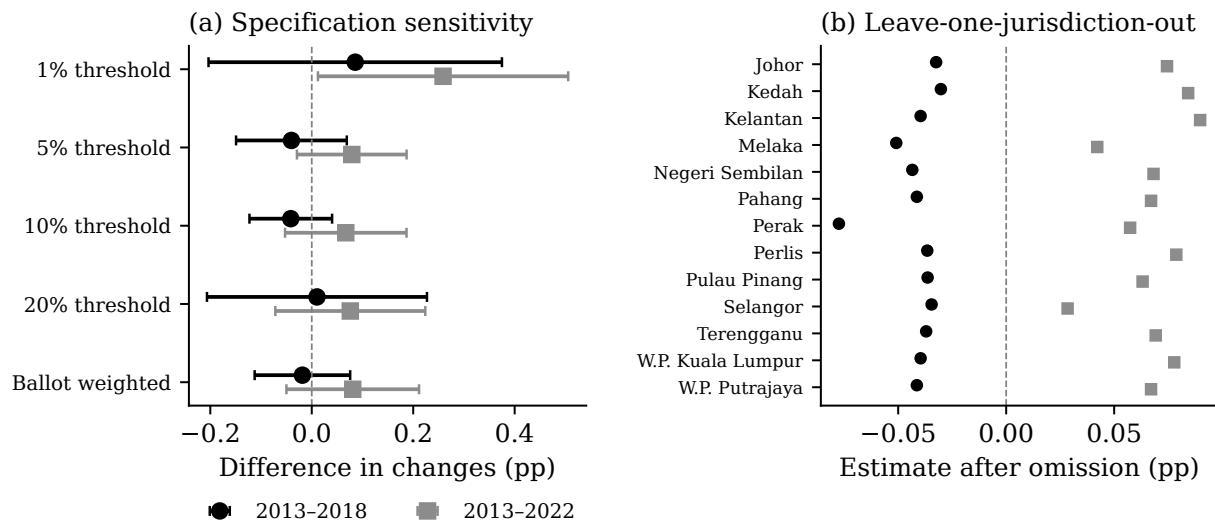


Figure 8. Sensitivity of the Malaysian difference-in-changes estimates. Panel (a) shows 95% confidence intervals across turnover thresholds and the 2013-ballot-weighted specification. Panel (b) shows point estimates after omitting each jurisdiction in turn; these are influence diagnostics, not additional independent tests.

2011 governed the elections of 2014, 2018, and 2022, and the map has been examined closely for partisan bias. Tanács-Mandák and Horváth (2025) show that the governing party's 2010 vote share predicts smaller constituency electorates under the 2011 map, a relationship absent under the map it replaced. ODIHR's observation mission for the 2022 election reported that 25 of the 106 constituencies deviated from the average electorate by more than 10 per cent, 18 by more than 15 per cent, and seven by more than 20 per cent, with a maximum deviation of 33 per cent (OSCE/ODIHR, 2022). The Roma dimension of the map is politically as well as informationally consequential: at the constituency level, the correlation between the Roma population share and the governing party's vote rose from 0.07 in 2014 to 0.58 in 2022 (Kovalcsik and Bódi, 2023).

Act LXXIX of 2024 then redrew the map for the election of 12 April 2026. ODIHR's assessment of preparations for that election reports that 39 single-member constituencies were redrawn. Budapest lost two seats, moving from 18 to 16, while Pest County gained two, moving from 12 to 14 (OSCE/ODIHR, 2026; Venice Commission, 2025). This establishes the institutional change needed for a before-and-after geographical comparison, but the count alone does not indicate which settlements lost information congruence.

The Hungarian design is specified on the same basis as the U.S. design. The units are the country's 3,155 settlements (Hungarian Central Statistical Office, 2026), with Budapest's districts and the larger cities divided at the finest published geography where a constituency line crosses them. Constituency and polling-station results for all four elections, 2014, 2018, 2022, and 2026, are published by the National Election Office (National Election Office, n.d.), and the 2022 census supplies settlement populations and Roma population shares (Hungarian Central Statistical Office, n.d.). The county component of the index uses the 19 counties and Budapest, within which constituencies must lie by law (Venice Commission, 2025), so that fragmentation in Hungary occurs where constituency lines divide settlements and the districts of the capital rather than counties. The before-and-after comparison contrasts settlements whose congruence fell under the 2024 map with settlements whose constituency assignment did not change, with the Roma share as the group dimension. Roma population shares require particular care because the observed category is based on reported identity, and its correspondence to the population reached by a given information source cannot be assumed. A geographical exposure measure should preserve the census definition while avoiding the inference that every resident of a Roma-dense settlement has the same media experience.

Hungarian media research identifies an additional mechanism that can operate independently of district

geometry. Szeidl and Szucs (2021) examine media capture through exchanges of favors involving advertising and political connections. Such relationships can alter the availability and character of reporting even when audience boundaries remain stable. This means that high geographical congruence may coexist with limited independent information about the relevant representative.

The minority-media case is more specific. Drawing on interviews and the histories of Rádió C and Dikh TV, Bleyer-Simon and Benedek (2025) examine the capture and institutional vulnerability of Romani ethnic media. Their evidence concerns ownership, funding, and the representation of minority voices. It identifies why a numerically small audience may face distinctive information constraints, while leaving open whether the 2024 constituency changes worsened those constraints. A settlement-level design would need to connect boundary exposure to independently measured coverage or knowledge before attributing a change to redistricting.

These findings imply that a comparative congruence model should include both geographical alignment and the conditions under which outlets operate. In Hungary, a measure of the share of a media audience contained in a constituency would be incomplete without evidence about the outlet's capacity to provide independent local political information. The same conceptual point applies elsewhere, although the institutional sources of constraint may differ.

6.2 Ethnicity, urban geography, and the Malaysian comparison

Malaysian delimitation is among the most studied cases of unequal constituency size. Ostwald (2013) shows how malapportionment shaped the 2013 result, in which the governing coalition lost the popular vote and retained a majority of the 222 seats, and Wong (2018) documents the delimitation process and the uses to which it has been put. For the 2018 boundaries, Ostwald (2019) reports that the largest constituency held roughly 9.1 times the electors of the smallest, and that the number of mixed peninsular constituencies, those between 40 and 60 per cent Malay, fell from 48 at the 2013 election to 33 at the 2018 election. A redelineation that sorts constituencies by ethnicity in this way changes which communities are divided, and the rejected-ballot comparison in Section 5 is the first outcome against which that change is measured here.

Malaysia also introduces a different set of measurement problems. Administrative districts, parliamentary constituencies, and the audiences of different language media do not form a single nested hierarchy. The administrative component of the index uses the 160 administrative districts for which the Department of Statistics publishes population by ethnic group (Department of Statistics Malaysia, n.d.a), and the parliamentary tables provide a separate geographical denominator (Department of Statistics Malaysia, n.d.b); joining the two requires an explicit geographical crosswalk. A constituency's ethnic composition cannot be inferred reliably from a similarly named administrative district, and a language-specific audience cannot be treated as identical to an ethnic category.

Research on Malaysian redelineation also cautions against reducing ethnic and geographical effects to one dimension. Brown (2005) finds that the 2002 exercise could reduce one form of ethnic bias while increasing disparities in constituency size. This illustrates how the consequences of a map depend on the outcome being evaluated. Population equality, ethnic representation, and information congruence may move differently even within the same boundary revision.

The present Malaysian analysis does not contain a harmonized ethnic-composition panel or voter-level measures of political knowledge. Its rejected-ballot estimates therefore cannot determine whether non-Malay-majority constituencies experienced a larger informational cost. Rejected ballots may reflect ballot design, voting procedures, intentional protest, or difficulty selecting a candidate, in addition to any informational effect. The weak aggregate association is compatible with several mechanisms and cannot resolve their relative importance.

A stronger test would link polling districts across the boundary change, recover the number and composition of residents reassigned, and distinguish candidate knowledge from ballot validity. Information about language-specific local coverage would make it possible to test whether a newly divided audience received less relevant reporting. This would preserve the central theory while measuring the channel more directly than constituency area turnover permits.

7 Ensemble benchmarks and a longitudinal design

7.1 The appropriate reference distribution

An enacted map may fragment communities partly because population equality and contiguity require some divisions. A neutral-plan ensemble offers a way to compare the enacted arrangement with alternative plans drawn under explicit constraints. The relevant benchmark is conditional on those constraints; it is not a universal distribution of all possible maps.

The ALARM project provides 5,000 simulated congressional plans for each of 44 multi-district states, for a total of 220,000 plans (McCartan et al., 2022; ALARM Project, n.d.). Those 44 states contain 429 of the 435 congressional districts; the six remaining states elect a single member at large, so every resident of them has a congruence of one by construction. The Census Bureau's 2020 tally lists 84,414 census tracts for the 50 states and the District of Columbia, of which 82,946 lie in the 44 states (Appendix A), and Nielsen's 210 designated market areas cover the continental United States, Hawaii, and parts of Alaska (U.S. Census Bureau, 2022; Nielsen, n.d.). These figures establish the scale of the U.S. reference resource. They do not make 220,000 plans independent observations of individual political knowledge. Analyses of ALARM plans demonstrate how ensembles can separate the consequences of geography from the consequences of a particular district arrangement (Kenny et al., 2023). Sequential Monte Carlo methods provide one approach to sampling plans subject to population and geographical constraints (McCartan and Imai, 2023).

For unit u in a state with B reference plans, an interpretable exposure contrast is

$$D_u = I_{u,p^{\text{enacted}}} - \text{median}_{b=1,\dots,B} I_{u,p^{(b)}}. \quad (9)$$

A negative value means that the enacted district gives the unit lower congruence than the median reference plan. An accompanying empirical rank is

$$Q_u = \frac{1 + \sum_{b=1}^B \mathbf{1}\{I_{u,p^{(b)}} \leq I_{u,p^{\text{enacted}}}\}}{B + 1}. \quad (10)$$

This is a descriptive position within the specified ensemble. It should not automatically be interpreted as a randomization-test p value, especially when the enacted plan was not sampled from the reference process and ensemble draws are dependent.

The group comparison should be computed plan by plan. For example, calculate the difference in population-weighted congruence between a minority population and a comparison population within each reference plan, then locate the enacted difference in that distribution. Computing a separate percentile for every tract and treating the tract percentiles as independent would understate their shared dependence on the same district plan. State-level aggregation also prevents states with many small tracts from receiving disproportionate weight merely because of their census geography.

7.2 Constraint choices and geographical harmonization

Recombination methods implemented in GerryChain provide a practical basis for reference plans outside the United States (DeFord et al., 2021; MGGG Redistricting Lab, n.d.). A reproducible application requires a

connected adjacency graph, a specified population target and tolerance, the number of seats, an initial plan, and explicit rules governing administrative divisions and protected communities. The reference distribution depends on these choices and on the sampling algorithm. Generating many plans cannot compensate for an incorrectly constructed graph or an inappropriate population constraint.

For Hungary, the graph could use settlements where they are sufficiently small, with finer units for settlements that must be divided. For Malaysia, polling districts or another sufficiently fine electoral geography would be preferable to large administrative districts that cannot be reassigned flexibly while maintaining population balance. Islands, enclaves, and permitted connections require explicit treatment. Every generated plan should be checked for seat count, population balance, contiguity, and compliance with the selected constraints.

Diagnostics should report the number of independent runs, acceptance behavior where applicable, variation in district assignments, and stability of the congruence summaries across runs. Agreement in a single scalar summary is insufficient if large portions of the map remain unchanged. Sensitivity ensembles should vary defensible assumptions about administrative splits and compactness. If a minority population appears disadvantaged only under one unusually restrictive reference specification, that dependence is part of the substantive result.

A common population baseline is necessary when old, enacted, and simulated maps are compared. Otherwise, a measured change in congruence can combine a boundary effect with population movement. The U.S. Census Bureau's geographical products and American Community Survey documentation support the construction of consistent small-area variables (U.S. Census Bureau, n.d.b); ACS table definitions specify the population variables (U.S. Census Bureau, n.d.a). Nielsen's DMA documentation identifies a relevant U.S. market geography, although audience boundaries and data vintages must be recorded explicitly (Nielsen, n.d.). Equivalent media units in Hungary and Malaysia need a country-specific justification rather than an assumed correspondence to U.S. DMAs.

7.3 Knowledge outcomes and difference-in-differences

The longitudinal knowledge model should use consistent questions, response coding, and geographical identifiers across survey waves. The Cooperative Election Study common-content releases for 2020, 2022, and 2024 contain about 61,000, 60,000, and 60,000 respondents, roughly 181,000 in all, and each carries congressional-district and county identifiers (Cooperative Election Study, 2021, 2023, 2025; Kuriwaki, 2025). Each wave asks respondents whether they have heard of their House representative and, if so, to which party he or she belongs; correct identification of the party is the knowledge outcome of the design. The combined respondent total is not automatically the sample size of a representative-party model: the relevant question must exist in each wave, apply to the same office, and link to the correct pre- or post-redistricting representative. Respondents lacking the required geography or outcome contribute no identified observation to that model.

Let K_{it} denote a harmonized knowledge outcome and $L_{it} = -\Delta I_{it}$ the loss of congruence between maps. A continuous-exposure specification is

$$K_{it} = \alpha_{it} + \lambda_t + \beta(L_{it} \times \text{Post}_t) + \theta' \mathbf{Z}_{it} + \varepsilon_{it}. \quad (11)$$

If the survey consists of repeated cross-sections rather than a respondent panel, α_{it} must refer to a consistently observed area or another valid repeated unit; respondent fixed effects cannot be introduced by treating different people as the same panel member. Survey weights and the geographical level of assignment should inform both estimation and uncertainty.

Interactions with baseline race, ethnicity, income, and English proficiency can test heterogeneity, but

their interpretation depends on whether the moderator is measured for the respondent or the area. A model with area minority share estimates contextual heterogeneity. It does not establish that minority respondents experience the estimated effect. Baseline values are preferable to contemporaneous measures that may themselves change through migration or the boundary revision.

Modern difference-in-differences methods make the assumptions behind such comparisons explicit (Callaway and Sant'Anna, 2021a). With continuous treatment, a slope across different exposure levels may require assumptions stronger than parallel trends between a treated and an untreated group (Callaway et al., 2021b). Multiple pre-periods should be used to inspect trends and assess sensitivity to departures from them (Rambachan and Roth, 2023). A statistically insignificant pretrend coefficient does not demonstrate that all relevant violations are absent.

The units of the coefficient also deserve explicit attention. If knowledge is coded from zero to one, a one-percentage-point decrease associated with a 0.10 decline in congruence corresponds to a slope of 0.10 with respect to congruence, or -0.10 with respect to congruence loss. This conversion prevents confusion between a percentage-point change in knowledge and a percentage change relative to the baseline. The precinct and Malaysian estimates reported here do not estimate that knowledge slope.

8 Discussion

The strongest new empirical finding is the positive association between U.S. precinct reassignment and House roll-off within a common district and election. It survives adjustment for population composition and weighting by precinct size. The evidence is consistent with the proposition that maintaining an established relationship with an incumbent reduces the informational cost of completing the local contest. Published studies of knowledge and local coverage supply independent support for the broader mechanism through which electoral geography can shape what citizens learn.

The distributional argument requires two further distinctions. An unequal exposure to fragmented boundaries is different from an unequal response to the same exposure. A community may bear a greater total burden because it is more frequently divided, even if its response per unit of fragmentation is no larger. Conversely, equal geographical exposure may produce unequal consequences where access to independent local news differs. The inconclusive composition interaction in the precinct analysis does not rule out unequal exposure, and the published evidence on racial differences in overlap does not establish a stronger individual knowledge response.

The Malaysian findings show why participation outcomes cannot be treated as interchangeable. A voter may know little about a constituency representative yet cast a valid vote using a party cue. A voter with substantial political knowledge may intentionally cast an invalid ballot. Boundary disruption can also change campaign organization or polling arrangements without changing the ability to identify a representative. These pathways make ballot rejection a useful but indirect outcome for the information hypothesis.

The Hungarian evidence adds an institutional qualification. Audience alignment can improve the economic relevance of local reporting, but it does not guarantee editorial independence or minority access. If a local outlet's incentives are shaped by ownership or political advertising, greater congruence may not yield more informative coverage. A comparative theory should therefore model geographical alignment and the organization of media provision together, rather than regard an overlap index as sufficient across all political settings.

Several limitations define the scope of the findings. The public U.S. file lacks the full coverage and covariates of the published original study. Its fixed effects do not remove unobserved precinct-level selection into reassignment. The Malaysian panel follows constituency lineages whose populations may change when

territory changes; its treatment is measured by area, and its inference relies on only 13 jurisdiction clusters. The comparative discussion does not estimate a new Hungarian settlement-level effect. These limits identify the specific observations needed to test the full argument, without negating the positive evidence about reassignment or the established literature on local information.

The practical research priority is a linked design that observes population-based congruence, exposure to relevant coverage, knowledge, and participation on consistent geographical units. Ensemble comparisons can then address whether a community's exposure is unusual relative to feasible alternatives, while longitudinal outcomes address whether the exposure matters. Each component answers a different question, and the credibility of the combined argument depends on preserving that distinction.

9 Conclusion

District boundaries organize access to political information as well as electoral representation. The U.S. reanalysis finds a positive association between incumbent reassignment and House roll-off, while the Malaysian comparison yields small and uncertain changes in rejected-ballot rates. Published research independently connects congruence to coverage and candidate knowledge, and documents unequal exposure to fragmented representation.

The Information Congruence Index measures population overlap with administrative communities and media audiences. The ensemble-based design specified here, with the ALARM plans for the United States and GerryChain reference plans for Hungary and Malaysia, is the direct test of whether district lines cut marginalized communities off from local political information beyond what neutral alternatives would produce. Comparisons with feasible alternative maps, combined with direct measures of political learning, are what will settle the size of the informational costs of electoral boundaries.

A Reproducibility and data construction

The accompanying source package contains the analysis script, the input data used for estimation, machine-readable results, vector figures, an executed notebook, and the LaTeX manuscript. Input-file checksums and package versions are recorded in the results directory. The statistical tables and figures can be regenerated by running `code/analyze_extended.py` from the package. The notebook displays the same stored results and provides the commands required to rerun the analysis.

A.1 U.S. sample construction

The source file is `vtds_merged.dta`, available from the data link associated with Hayes and McKee (2012) on the author's research page. The package retains the original downloaded bytes under `hayes_mckee_2012.dta`. The all-zero-row test is applied across the entire record, so a genuine zero on the outcome alone does not lead to exclusion. The analysis then requires `contest=1`, `open=0`, positive `vtdsize`, and nonmissing values for the modeled variables.

Composition controls are included as continuous variables. The interaction specification contains main effects for reassignment and all three composition variables, their three reassignment interactions, and district–election effects. Conditional contrasts use the full estimated covariance matrix, including covariances among main and interaction terms. Calculating the uncertainty from the interaction coefficient alone would give incorrect intervals for those contrasts.

Table 6. U.S. variables used in the reanalysis

Source variable	Role in the analysis	Treatment
rolloff	Dependent variable	Retained in its supplied percentage-point units
redrawn	Reassignment indicator	Binary exposure
blackper	Population composition	Divided by 10 for interaction interpretation
hispper	Population composition	Divided by 10
othper	Population composition	Divided by 10
stateyear	Election stratum	Categorical effect; original numeric codes retained
district	District identifier	Interacted with stratum to construct 102 cells
vtdsize	Precinct size	Positive-value restriction; weight in column 4
contest, open	Sample restrictions	Contested races with returning incumbents

A.2 Malaysian geographical correspondence

The geographical inputs contain 166 constituencies in each map, including Labuan. The analysis excludes Labuan after validating the geometries and computing the cross-map correspondence. Intersection and union areas are calculated after projection rather than directly from longitude and latitude. The maximum-overlap check operates within jurisdiction and requires a one-to-one assignment, preventing two old constituencies from being matched to the same new constituency.

A matched constituency lineage should not be interpreted as a fixed set of residents. The comparison follows corresponding seats as their territories change. Consequently, the difference-in-changes coefficient combines any behavioral response with changes in the composition of the population voting in that seat. A population-preserving study would instead construct stable intersections of old and new boundaries and allocate election or survey outcomes to those units where the source data permit it.

The baseline cutoff places 51 lineages in the higher-turnover group and 114 in the lower-turnover group. At the 1%, 5%, and 20% cutoffs, the higher-turnover counts are 158, 89, and 25, respectively. These counts explain why a very low cutoff is a demanding robustness check: the comparison group becomes small and geographically selective. All estimated threshold models are retained in the machine-readable results rather than filtered by statistical significance.

A.3 A consistent geographical denominator

The tract count used in a national reference dataset must correspond to the states included in the districting analysis. The Census Bureau's 2020 entity tally lists 84,414 tracts for the 50 states and the District of Columbia, whereas the 44 multi-district states contain 82,946 tracts, including water-only entities (U.S. Census Bureau, 2022). These totals describe geographical inventories, not the number of nonempty, successfully linked observations in an outcome model. Population filters, map intersections, and survey matching will change the analytic count and should be documented separately.

Likewise, the number of congressional districts, ensemble plans, media markets, survey respondents, and election records are different dimensions of a study. None should be presented as a substitute for the sample size supporting a particular coefficient. For each model, the replication files should identify the observational unit, the usable observations, the geographical clusters, and the effective source of variation.

B Additional Malaysian estimates

Table 7. Threshold sensitivity in the Malaysian panel

Turnover threshold	Higher-turnover lineages	2013–2018	2013–2022
Greater than 1%	158	0.086	0.259
95% CI		[−0.204, 0.375]	[0.012, 0.506]
Greater than 5%	89	−0.040	0.079
95% CI		[−0.149, 0.069]	[−0.029, 0.187]
Greater than 10%	51	−0.041	0.067
95% CI		[−0.123, 0.040]	[−0.053, 0.187]
Greater than 20%	25	0.010	0.076
95% CI		[−0.207, 0.227]	[−0.072, 0.224]

Each post-period model uses 165 constituency lineages and jurisdiction fixed effects. Coefficients are percentage-point differences in changes in rejected-ballot rates. Intervals use clustered standard errors and t_{12} critical values.

The main continuous models use turnover divided by 0.10, which makes their coefficients comparable to a ten-percentage-point increase in the geographical exposure. This scaling does not transform territorial turnover into population congruence. The weighted specification uses baseline ballots rather than post-redelineation ballots to avoid allowing a post-treatment turnout change to determine the observation's weight. Leave-one-jurisdiction-out estimates are reported as point-estimate diagnostics because their purpose is to assess influence, not to supply an alternative set of significance tests.

References

- Algorithm-Assisted Redistricting Methodology Project. (n.d.). 50-state redistricting simulations and frequently asked questions. <https://alarm-redist.org/fifty-states/about/>
- Balles, P., Matter, U., & Stutzer, A. (2023). Television market size and political accountability in the U.S. House of Representatives. *European Journal of Political Economy*, 80, 102459. <https://doi.org/10.1016/j.ejpoleco.2023.102459>
- Bleyer-Simon, K., & Benedek, K. (2025). State capture of Romani ethnic media in Hungary. *European Journal of Communication*, 40(1), 23–36. <https://doi.org/10.1177/02673231241282509>
- Brown, G. K. (2005). Playing the (non)ethnic card: Ethnicity and the electoral system in Malaysia. *Ethnopolitics*, 4(4), 429–445. <https://doi.org/10.1080/17449050500348675>
- Callaway, B., & Sant'Anna, P. H. C. (2021). Difference-in-differences with multiple time periods. *Journal of Econometrics*, 225(2), 200–230. <https://doi.org/10.1016/j.jeconom.2020.12.001>
- Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. C. (2021). Difference-in-differences with a continuous treatment [Working paper]. arXiv. <https://doi.org/10.48550/arXiv.2107.02637>
- Cooperative Election Study. (2021). Cooperative Election Study Common Content, 2020 [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/E9N6PH>
- Cooperative Election Study. (2023). Cooperative Election Study Common Content, 2022 [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/PR4L8P>
- Cooperative Election Study. (2025). Cooperative Election Study Common Content, 2024 [Data set]. Harvard Dataverse. <https://doi.org/10.7910/DVN/X11EP6>
- Curiel, J. A., & Steelman, T. (2018). Redistricting out representation: Democratic harms in splitting zip codes. *Election Law Journal*, 17(4), 328–353. <https://doi.org/10.1089/elj.2018.0528>
- Darr, J. P., Hitt, M. P., & Dunaway, J. L. (2018). Newspaper closures polarize voting behavior. *Journal of Communication*, 68(6), 1007–1028. <https://doi.org/10.1093/joc/jqy051>
- DeFord, D., Duchin, M., & Solomon, J. (2021). Recombination: A family of Markov chains for redistricting. *Harvard Data Science Review*, 3(1). <https://hdsr.mitpress.mit.edu/pub/1ds8ptxu/release/5>

- Department of Statistics Malaysia. (n.d.-a). Population table: Administrative districts [Data catalogue]. https://data.gov.my/data-catalogue/population_district
- Department of Statistics Malaysia. (n.d.-b). Population table: Parliamentary constituencies [Data catalogue]. https://open.dosm.gov.my/data-catalogue/population_parlimen
- ElectionData.MY. (n.d.-a). 2003 delimitation of Peninsular Malaysia (Parliament) [Data set]. <https://electiondata.my/data-catalogue/peninsular-2003-parlimen/>
- ElectionData.MY. (n.d.-b). 2018 delimitation of Peninsular Malaysia (Parliament) [Data set]. <https://electiondata.my/data-catalogue/peninsular-2018-parlimen/>
- ElectionData.MY. (n.d.-c). Parliamentary elections (PRU): Statistics [Data set]. <https://electiondata.my/data-catalogue/headline-stats-federal/>
- Hayes, D., & Lawless, J. L. (2015). As local news goes, so goes citizen engagement: Media, knowledge, and participation in US House elections. *The Journal of Politics*, 77(2), 447–462. <https://doi.org/10.1086/679749>
- Hayes, D., & Lawless, J. L. (2018). The decline of local news and its effects: New evidence from longitudinal data. *The Journal of Politics*, 80(1), 332–336. <https://doi.org/10.1086/694105>
- Hayes, D., & McKee, S. C. (2009). The participatory effects of redistricting. *American Journal of Political Science*, 53(4), 1006–1023. <https://doi.org/10.1111/j.1540-5907.2009.00413.x>
- Hayes, D., & McKee, S. C. (2012). The intersection of redistricting, race, and participation. *American Journal of Political Science*, 56(1), 115–130. <https://doi.org/10.1111/j.1540-5907.2011.00546.x>
- Hood, M. V., III, & McKee, S. C. (2010). Stranger danger: Redistricting, incumbent recognition, and vote choice. *Social Science Quarterly*, 91(2), 344–358. <https://doi.org/10.1111/j.1540-6237.2010.00696.x>
- Huber, G. A., & Tucker, P. D. (2024). House members on the news: Local television news coverage of incumbents. *British Journal of Political Science*, 54(2), 503–513. <https://doi.org/10.1017/S0007123423000480>
- Hungarian Central Statistical Office. (n.d.-b). Key data on settlements: Census 2022. <https://nepszamlalas2022.ksh.hu/en/results/visualizations/key-data-on-settlements>
- Hungarian Central Statistical Office. (2026). Number and population of settlements by legal status, 1 January 2026 [STADAT table 8.1.2.2]. https://www.ksh.hu/stadat_files/fo/hu/fo10007.html
- Jerit, J., Barabas, J., & Bolsen, T. (2006). Citizens, knowledge, and the information environment. *American Journal of Political Science*, 50(2), 266–282. <https://doi.org/10.1111/j.1540-5907.2006.00183.x>
- Kenny, C. T., McCartan, C., Simko, T., Kuriwaki, S., & Imai, K. (2023). Widespread partisan gerrymandering mostly cancels nationally, but reduces electoral competition. *Proceedings of the National Academy of Sciences*, 120(25), e2217322120. <https://doi.org/10.1073/pnas.2217322120>
- Kovalcsik, T., & Bódi, M. (2023). Geographical realignment of the Hungarian voting behaviour between 2014 and 2022. *Modern Geográfia*, 18(1), 59–77. <https://doi.org/10.15170/MG.2023.18.01.04>
- Kuriwaki, S. (2025). Guide to the cumulative common content of the Cooperative (Congressional) Election Study, 2006–2024. Cooperative Election Study. https://csmweb-prod-02.ist.berkeley.edu/sdaweb/docs/ces-cumulative-2024-v10/DOC/guide_cumulative_2006-2024.pdf
- Lawrence, C. N., & Huffmon, S. H. (2015). Keeping up with the congressmen: Evaluating constituents' awareness of redistricting. *Social Science Quarterly*, 96(1), 65–75. <https://doi.org/10.1111/ssqu.12130>
- McCartan, C., & Imai, K. (2023). Sequential Monte Carlo for sampling balanced and compact redistricting plans. *The Annals of Applied Statistics*, 17(4), 3300–3323. <https://imai.fas.harvard.edu/research/smcredist/>
- McCartan, C., Kenny, C. T., Simko, T., Garcia, G., III, Wang, K., Wu, M., Kuriwaki, S., & Imai, K. (2022). Simulated redistricting plans for the analysis and evaluation of redistricting in the United States. *Scientific Data*, 9, 689. <https://doi.org/10.1038/s41597-022-01808-2>
- MGGG Redistricting Lab. (n.d.). GerryChain documentation: Running a chain with ReCom [Software documentation]. <https://mggg.github.io/GerryChain/user/recom/>
- Moskowitz, D. J. (2021). Local news, information, and the nationalization of U.S. elections. *American Political Science Review*, 115(1), 114–129. <https://doi.org/10.1017/S0003055420000829>
- Myers, A. C. W. (2026). Press coverage and accountability in state legislatures. *American Political Science Review*, 120(1), 207–225. <https://doi.org/10.1017/S000305542500022X>
- National Election Office of Hungary. (n.d.). Parliamentary elections 2014, 2018, 2022, and 2026: Official results and downloadable data. <https://www.valasztas.hu/orszaggyulesi-valasztasok>
- Nielsen. (n.d.). ZIP code by DMA regions [Product documentation]. <https://www.nielsen.com/dma-regions/>

- Niven, D., Combs, B. H., Norwood, C., Rossiter, K. E., & Solimine, M. E. (2022). The boundaries of confusion: Gerrymandering and racial disparities in state House and congressional district line congruity. *Social Science Quarterly*, 103(6), 1507–1518. <https://doi.org/10.1111/ssqu.13213>
- Niven, D., & Solimine, M. E. (2022). Representing people and places: Castaway voters and the racial disparity in redistricting. *Election Law Journal: Rules, Politics, and Policy*, 21(2), 171–186. <https://doi.org/10.1089/elj.2021.0032>
- OSCE Office for Democratic Institutions and Human Rights. (2022). Hungary: Parliamentary elections and referendum, 3 April 2022. ODIHR election observation mission final report. Warsaw, 29 July 2022. <https://cdn.osce.org/sites/default/files/f/documents/2/6/523568.pdf>
- OSCE Office for Democratic Institutions and Human Rights. (2026). Hungary: Parliamentary elections, 12 April 2026. ODIHR needs assessment mission report, 15–18 December 2025. https://resources.osce.org/sites/default/files/documents/official_documents/2026/01/HUN%20parliamentary%202026_NAM%20report_21.01.2026.pdf
- Ostwald, K. (2013). How to win a lost election: Malapportionment and Malaysia's 2013 general election. *The Round Table*, 102(6), 521–532. <https://doi.org/10.1080/00358533.2013.857146>
- Ostwald, K. (2019). Electoral boundaries in Malaysia's 2018 election: Malapportionment, gerrymandering and UMNO's fall. In E. T. Gomez & M. N. Mohamed Osman (Eds.), *Malaysia's 14th general election and UMNO's fall: Intra-elite feuding in the pursuit of power* (pp. 86–110). Routledge. <https://doi.org/10.4324/9780429318375-5>
- Rambachan, A., & Roth, J. (2023). A more credible approach to parallel trends. *The Review of Economic Studies*, 90(5), 2555–2591. <https://doi.org/10.1093/restud/rdad018>
- Robinson, W. S. (1950). Ecological correlations and the behavior of individuals. *American Sociological Review*, 15(3), 351–357. <https://doi.org/10.2307/2087176>
- Snyder, J. M., Jr., & Strömberg, D. (2010). Press coverage and political accountability. *Journal of Political Economy*, 118(2), 355–408. <https://doi.org/10.1086/652903>
- Szeidl, A., & Szucs, F. (2021). Media capture through favor exchange. *Econometrica*, 89(1), 281–310. <https://doi.org/10.3982/ECTA15641>
- Tanács-Mandák, F., & Horváth, A. (2025). The “hacking” of a mixed electoral system: A case study of Hungary. *Public Choice*, 204, 75–99. <https://doi.org/10.1007/s11227-025-01296-z>
- Thevananthan, T. (2026). The Malaysian Election Corpus (MECo): Federal and state-level election results from 1955 to 2025. *Scientific Data*, 13, 190. <https://doi.org/10.1038/s41597-025-06502-7>
- Thevananthan, T., & Chacko, D. P. (2025). The Malaysian Election Corpus (MECo): Electoral maps and cartograms from 1954 to 2025 [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2512.24211>
- U.S. Census Bureau. (2022). 2020 Census geographic entity tallies by state [Data set]. https://www2.census.gov/geo/docs/maps-data/data/geo_tallies2020/2020talliesbystate.xlsx
- U.S. Census Bureau. (n.d.-a). 2020 American Community Survey five-year detailed tables: Group definitions [Data documentation]. <https://api.census.gov/data/2020/acs/acs5/groups.html>
- U.S. Census Bureau. (n.d.-b). 2020 TIGER/Line shapefiles [Data documentation]. <https://www.census.gov/geographies/mapping-files/2020/geo/tiger-line-file.html>
- European Commission for Democracy through Law (Venice Commission). (2025). Hungary: Opinion on Act LXXIX of 2024 amending certain laws relating to elections (CDL-AD(2025)018), adopted at the 143rd plenary session, 13–14 June 2025. <https://docs-venice.coe.int/api/Document?pdffile=CDL-AD%282025%29018-e>
- Wong, C. H. (2018). Constituency delimitation and electoral authoritarianism in Malaysia. *The Round Table*, 107(1), 67–80. <https://doi.org/10.1080/00358533.2018.1424075>
- Winburn, J., & Wagner, M. W. (2010). Carving voters out: Redistricting's influence on political information, turnout, and voting behavior. *Political Research Quarterly*, 63(2), 373–386. <https://doi.org/10.1177/1065912908330728>