Geographic Representativeness in Social Media Surveys

Evidence from the 2025 German Federal Election Study

Simon Munzert · Together with Cornelius Erfort, Thomas Gschwend, Elias Koch, Hannah Rajski, Lukas Stoetzer

DEMED Webinar Series · 19 May 2026

zweitstimme.org

DFG Deutsche Forschungsgemeinschaft

Motivation

The standard representativeness checklist

What we routinely benchmark

- **Age** — young people underrepresented in most modes
- **Gender** — usually close but not perfect
- **Education** — higher-educated overrepresented online

What we do about imbalances

- Report demographic distributions vs. population benchmarks
- Post-stratification / raking weights using census benchmarks
- Report design effects (DEFF), effective N
- Sensitivity analyses with and without weights

Pop benchmarking: Neundorf/Öztürk 2023, PLOS One

	Conversion	Traffic	Population Benchmark
United Kingdom			
Targeted categories			
Young (18–34)	7%***	4%***	29%
Middle-aged (35–54)	17%***	11%***	34%
No college	39%***	44%***	70%
Female	42%***	44%***	51%
Non-targeted categories			
Less interested in politics	24%***	24%***	43%
Not partisan	42%***	47%	48%
Total	n = 1,033	n = 756	
Turkey			
Targeted categories			
Young (18–34)	21%***	19%***	37%
Middle-aged (35–54)	38%	27%***	40%
No college	32%***	73%*	80%
Female	32%***	34%***	50%
Non-targeted categories			
Less interested in politics	33%***	54%	48%
Not partisan	42%***	56%***	32%
Total	n = 1,689	n = 203	
Spain			
Targeted categories			
Young (18–34)	10%***	5%***	22%
Middle-aged (35–54)	29%***	20%***	41%
No college	43%***	57%***	73%
Female	45%***	36%**	52%
Non-targeted categories			
Less interested in politics	17%***	28%***	60%
Not partisan	22%***	34%	41%
Total	n = 1,709	n = 92	
Czech Republic			
Targeted categories			
Young (18–34)	35%***	20%	25%
Middle-aged (35–54)	40%**	23%***	36%
No college	49%***	65%***	83%
Female	32%***	42%*	51%
Non-targeted categories			
Less interested in politics	24%***	51%***	82%
Not partisan	42%***	59%	61%
Total	n = 1,634	n = 193	

Note: Chi-square goodness of fit comparing Meta and population samples:
* = p≤0.05,
** = p≤0.01,
*** = p≤0.001.
The percentages refer to the proportion of respondents as a share of the overall sample in each campaign (column percentages). For example, 42% of respondents recruited through a conversion campaign in the UK are women (and conversely 58% are men). *Population benchmark*: We use data from the round 9 of European Social Survey (for UK, Spain, and CZ) and from the wave 5 of Comparative Study of Electoral Systems (for Turkey) for population benchmarks. We weighted population estimates from these surveys with weight variables provided in the original survey. We used *pspweight* variable for ESS survey and *E1010_2* variable for CSES survey. According to codebooks of these surveys, both of these weight variables were constructed based on demographic information.

https://doi.org/10.1371/journal.pone.0281243.t003

The standard representativeness checklist

Hertie School

What we routinely benchmark

- **Age** — young people underrepresented in most modes
- **Gender** — usually close but not perfect
- **Education** — higher-educated overrepresented online

What we do about imbalances

- Report demographic distributions vs. population benchmarks
- Post-stratification / raking weights using census benchmarks
- Report design effects (DEFF), effective N
- Sensitivity analyses with and without weights

Census benchmarking: this study

			Sample Share Compared to Benchmark (Census Data)		
				Comparing Sub-Samples	
	Group	Benchmark	Full Sample	Bilendi Sample	Social Media Sample
Age	15-24 y.o.	10.0%	6.5% (-3.6pp)	4.2% (-5.9pp)	8.6% (-1.5pp)
	25-34 y.o.	12.4%	15.9% (+3.5pp)	14.9% (+2.5pp)	16.8% (+4.4pp)
	35-44 y.o.	13.3%	19.6% (+6.3pp)	21.6% (+8.4pp)	17.7% (+4.4pp)
	45-54 y.o.	12.5%	17.8% (+5.2pp)	19.2% (+6.7pp)	16.4% (+3.9pp)
	55-64 y.o.	15.7%	25.9% (+10.2pp)	26.1% (+10.4pp)	25.7% (+10.0pp)
	65 y.o. or older	21.9%	14.4% (-7.5pp)	13.9% (-8.0pp)	14.8% (-7.1pp)
Gender	Female	50.6%	48.6% (-2.0pp)	51.9% (+1.3pp)	45.5% (-5.0pp)
	Male	49.4%	51.4% (+2.0pp)	48.1% (-1.3pp)	54.5% (+5.0pp)
Education	High	36.6%	61.6% (+25.0pp)	53.5% (+16.9pp)	69.0% (+32.4pp)
	Low	30.3%	7.4% (-22.9pp)	10.3% (-20.0pp)	4.8% (-25.6pp)
	Medium	29.5%	30.2% (+0.7pp)	35.7% (+6.2pp)	25.1% (-4.3pp)
	Still in school	3.6%	0.8% (-2.8pp)	0.5% (-3.1pp)	1.1% (-2.5pp)

Why geographic representativeness matters I

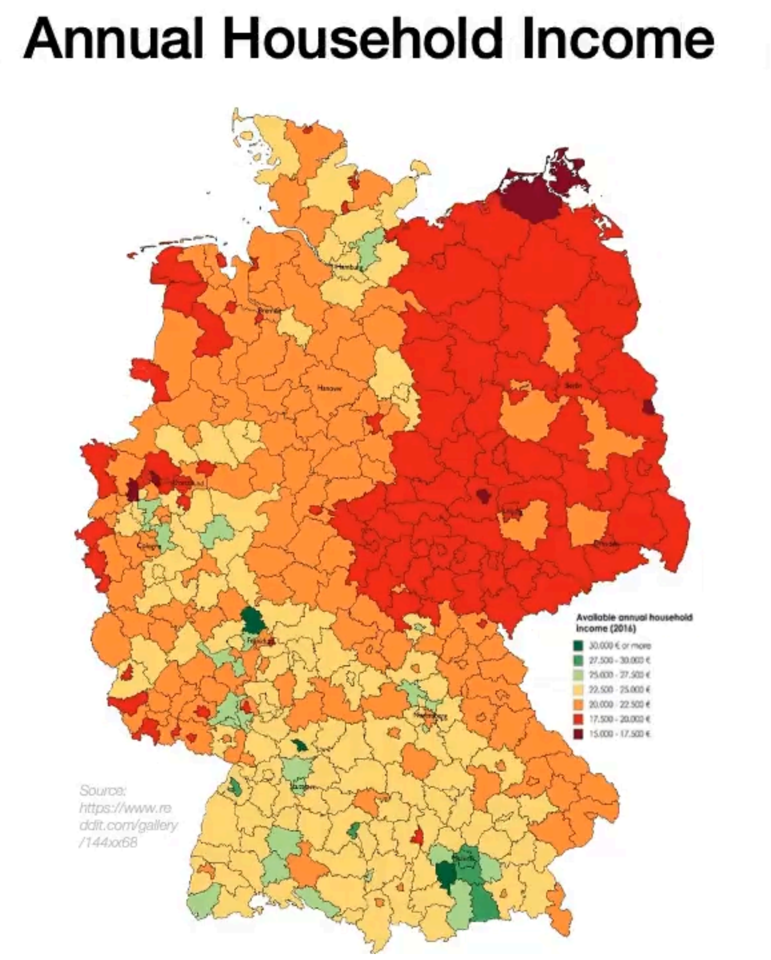
Hertie School

Many social phenomena are inherently geographic

- Urban-rural attitude divides
- Local media and political environment effects
- District-level election outcomes
- Spatial diffusion of information

Political processes are structured by geography

- **East-West divide in Germany**: AfD support, trust in institutions, economic grievances — not reducible to age/education
- **Urban-rural cleavage**: party system transformation across Western democracies
- **Electoral systems may amplify geography**: local representation can boost idiosyncracy of local opinion

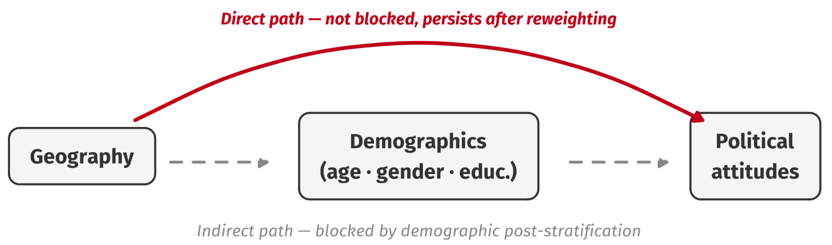


Source: Tomas Pueyo | ixtract

Why geographic representativeness matters II

Geographic bias can't be fully corrected demographically

- Respondents with identical demographics can differ on other variables if they live in different regions
- With insufficient geographic variation, area-level covariates are poorly estimated



The downstream consequences

For description/prediction

- MRP requires geographic variation to produce reliable small-area estimates (Gelman & Little 1997)
- Regional opinion estimates are biased toward over-represented areas
- Systematic under-coverage of certain regions inflates out-of-sample forecast uncertainty

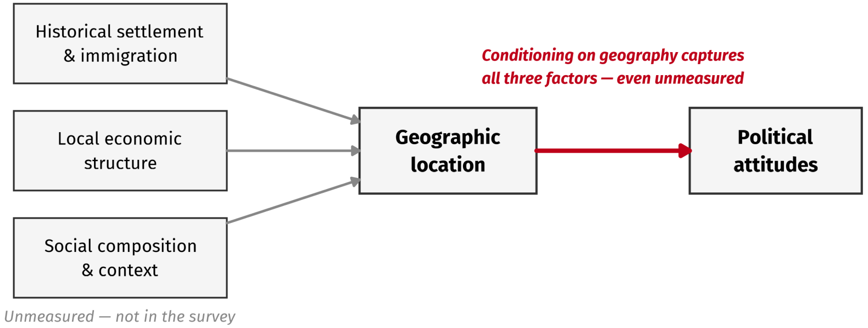
For causal inference

- Treatment effect heterogeneity by geography cannot be analyzed without geographic variation
- Regression coefficients on geographic moderators/confounders potentially biased when coverage is uneven

Why geographic representativeness matters III

Efficient balancing on unmeasured factors

Geography bundles many determinants of political attitudes that surveys never directly measure — historical settlement patterns, local economic structure, social composition. Conditioning on geography helps absorb all of them simultaneously.



Enabling substantive research

Political predispositions and behaviors **come in geographic clusters** (Agnew 2002; Johnston & Pattie 2006; Gelman 2008). Without geographic coverage, these patterns are invisible.

- **Historical settlement & immigration** — long-run regional political cultures (Elazar 1994)
- **Industrialization & urbanization** — structural sorting of populations (Rodden 2010)
- **Social composition** — neighborhood-level context effects (Hero 1998)
- **Local economic conditions** — place-based grievances and economic voting (Heppen 2003)

Geographic recruitment in classical survey research

Face-to-face and CAPI surveys

Traditional probability surveys use **geographic clustering** as a deliberate design feature to reduce fieldwork cost:

- **Primary Sampling Units (PSUs)**: municipalities, census tracts, or enumeration areas drawn first
- **Secondary/tertiary stages**: households and individuals within selected PSUs

Implications

- **Effective N**: nominal N / DEFF — often 50–70% of raw N in clustered designs
- Geographic coverage is **by design incomplete** — only selected PSUs are visited

`References:` Kish (1965); Groves et al. (2009)

Examples

- **ALLBUS / ISSP Germany**: stratified multi-stage cluster sample; 162 PSUs; geographic spread ensured by stratification but not full coverage
- **European Social Survey**: Probability sampling with PSU selection; clustering leads to DEFF ≈ 1.5–2.5
- **BHPS/SOEP**: Initial clustered recruitment; refreshment samples face same geographic constraints

The contrast with online/social media

Online convenience and opt-in panels **lack geographic probability selection** — geographic coverage depends entirely on platform penetration and targeting choices, not on a sampling frame.

`References:` Gabler & Häder (1997)

Four questions this talk addresses

Q1: Recruitment

How can we recruit geographically on social media?

What targeting capabilities do platforms offer? What are the trade-offs as we move below the country level?

Q2: Measurement

How do we measure geographic representativeness?

At which geographic level — and using which information source — should we assess coverage?

Q3: Predictors

What predicts geographic coverage?

Are there factors — political, structural, digital — that predict which areas are over- or under-represented?

(Q4: Downstream effects)

Does geographic coverage affect our estimates?

When does geographic imbalance translate into substantively meaningful bias in measured outcomes?

Geographically Targeted Recruitment on Social Media

What social media platforms offer for geographic targeting

Facebook/Instagram (↪ link for details)

- Country/State/Province/City/Zip codes/Addresses; set radius (depending on location type)

YouTube

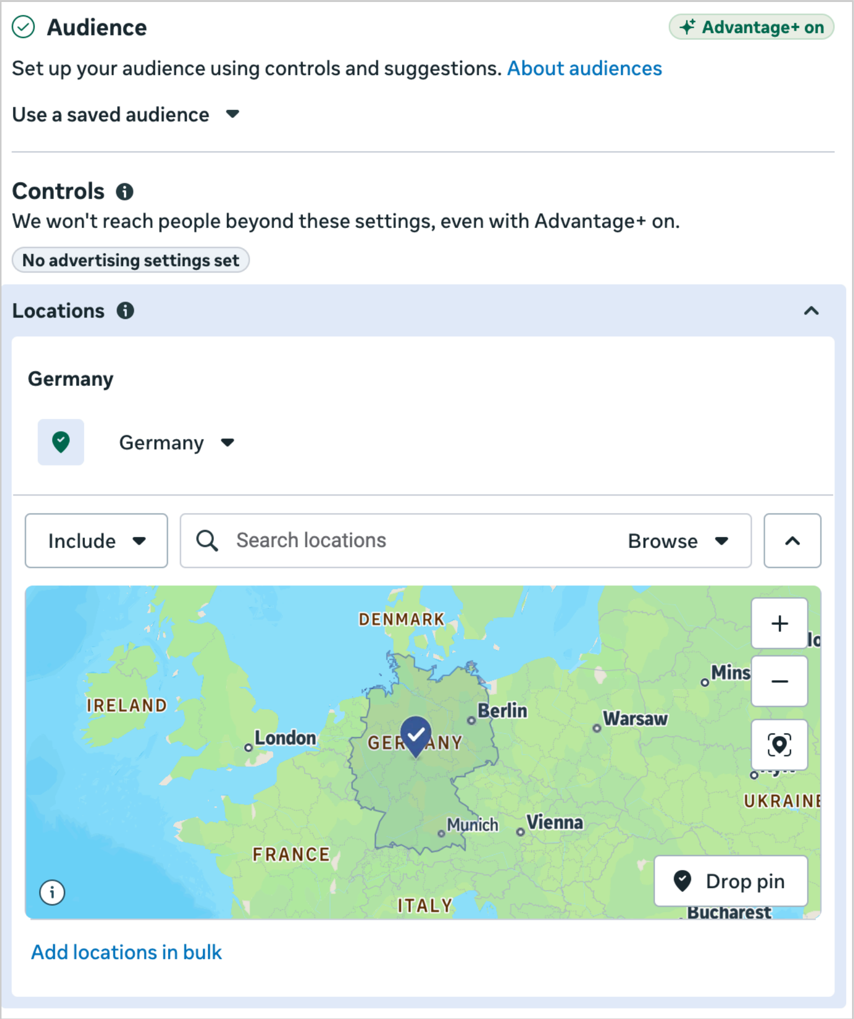
- Many location types and levels, availability varies by country; check the full table

TikTok

- Many location levels from country to zip code; "target up to 3,000 location selections at a time, per ad group"

X (formerly Twitter)

- Countries, regions, metros, cities, postal codes, radius (USA, JPN)



Meta's geographic targeting in practice

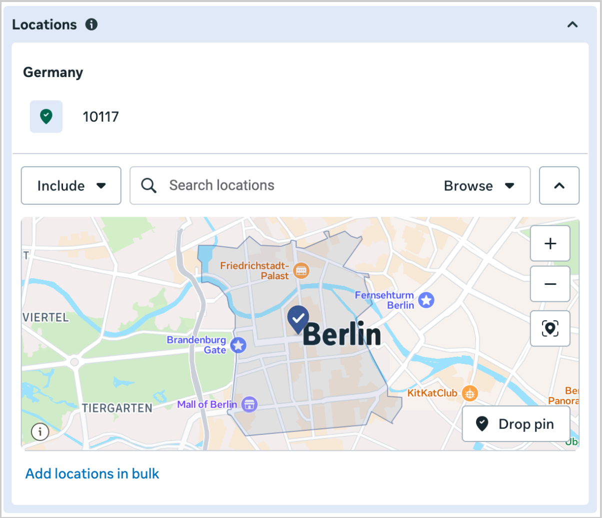
Campaign structure

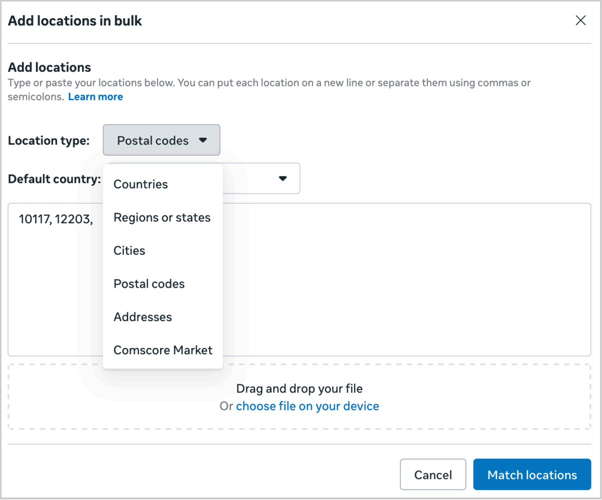
- Targeting is set at the **Ad Set** level; each Ad Set targets one or more geographic units
- **Location type**: living there vs. recently there vs. traveling
- **Reach estimates**: Meta shows estimated audience size before launch

Setting up multiple ad sets for varying locations manually in Ads Manager is tedious — Meta supports **bulk creation via CSV/spreadsheet upload**. You define each ad set as a row.

Caveats

- Even with state-level targeting, **IP geolocation ≠ actual location** — users in border areas, VPN users, or frequent travelers might get misassigned.
- Bid strategy: some areas (e.g., urban) might be more competitive (higher CPMs)





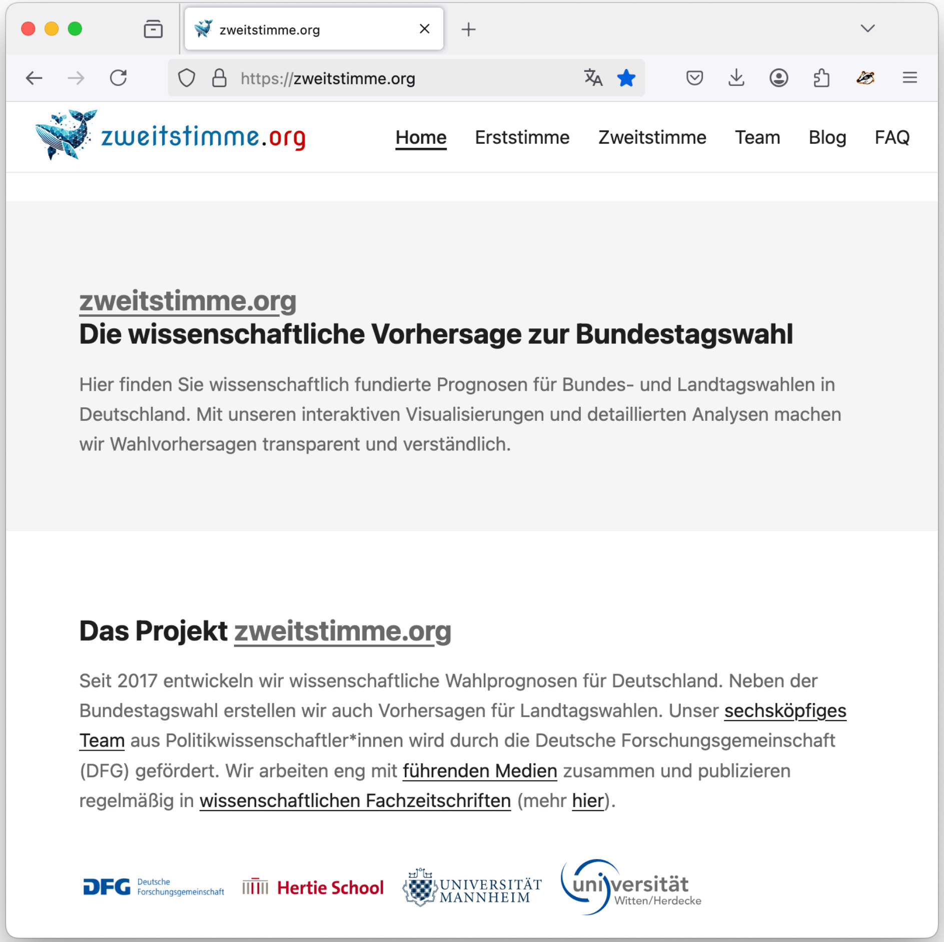
Case Study: 2025 German Federal Election Study

zweitstimme.org: forecasting the German federal election

Hertie School

The project

- Election forecasting and communication project for the 2025 German federal election (Bundestagswahl), funded by the German Science Foundation (DFG)
- Research on **citizen forecasting at the district level**
- Research on effective communication of poll results to the public



Study design: dual-mode pre-election survey

Hertie School

Context

- Fielded ahead of the **2025 German federal election**
- Two-wave design; hosted on Qualtrics
- Two modes of recruitment: (1) social media ads (N = 23,570); (2) online access panel (N = 21,318)

Access panel recruitment (Bilendi)

- No strict quotas; soft targets: >20 resp per electoral district (299), 50/50 gender balance

Social media recruitment

- Facebook + Instagram placements, 64 ad sets
- State (16) x age group (2) x gender (2)
- Budgets $\propto$ state-level eligible voter counts

Location

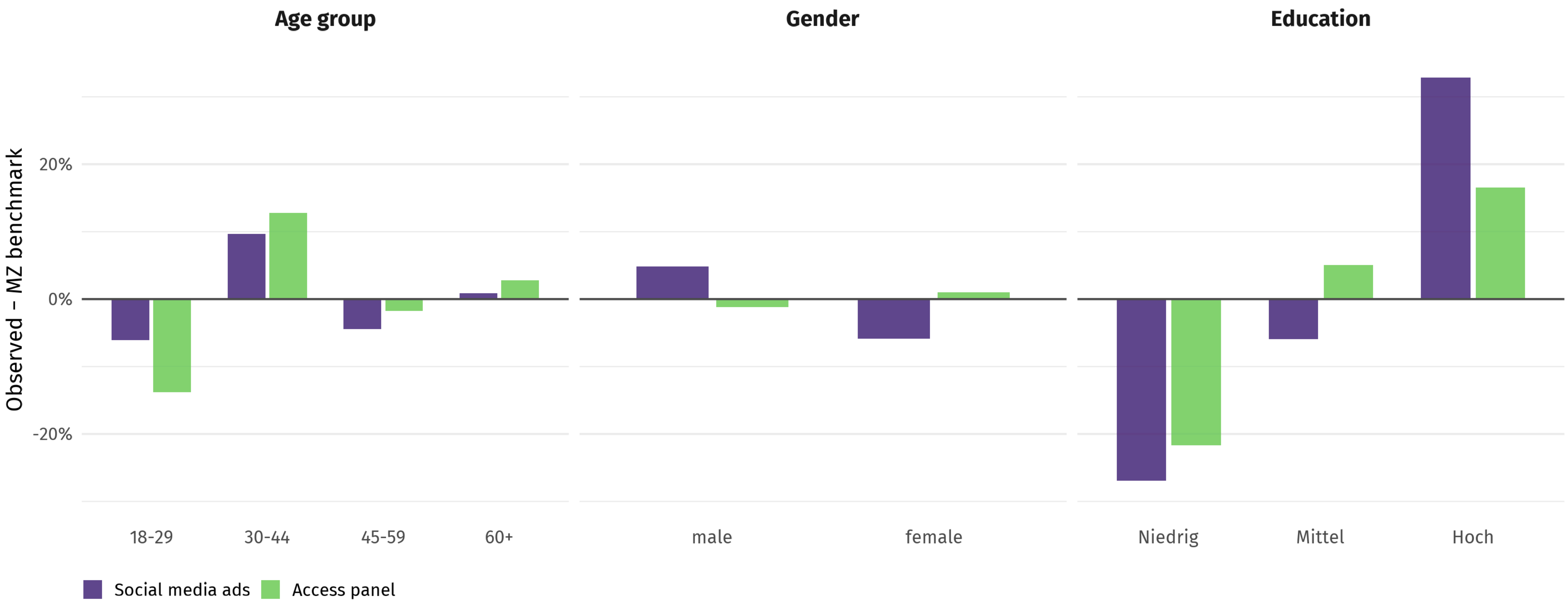
1. Targeted by **Bundesland**
2. Survey measures three geographic indicators:
   - **ZIP code** (free text, mapped to WK via official crosswalk)
   - **Self-reported Wahlkreis** (derived from ZIP code and drilldown dropdown)
   - **IP geolocation** (automatic via Qualtrics: lat/lon → spatial join → WK)

Sample composition vs. population benchmarks

Hertie School

Deviation from population benchmark

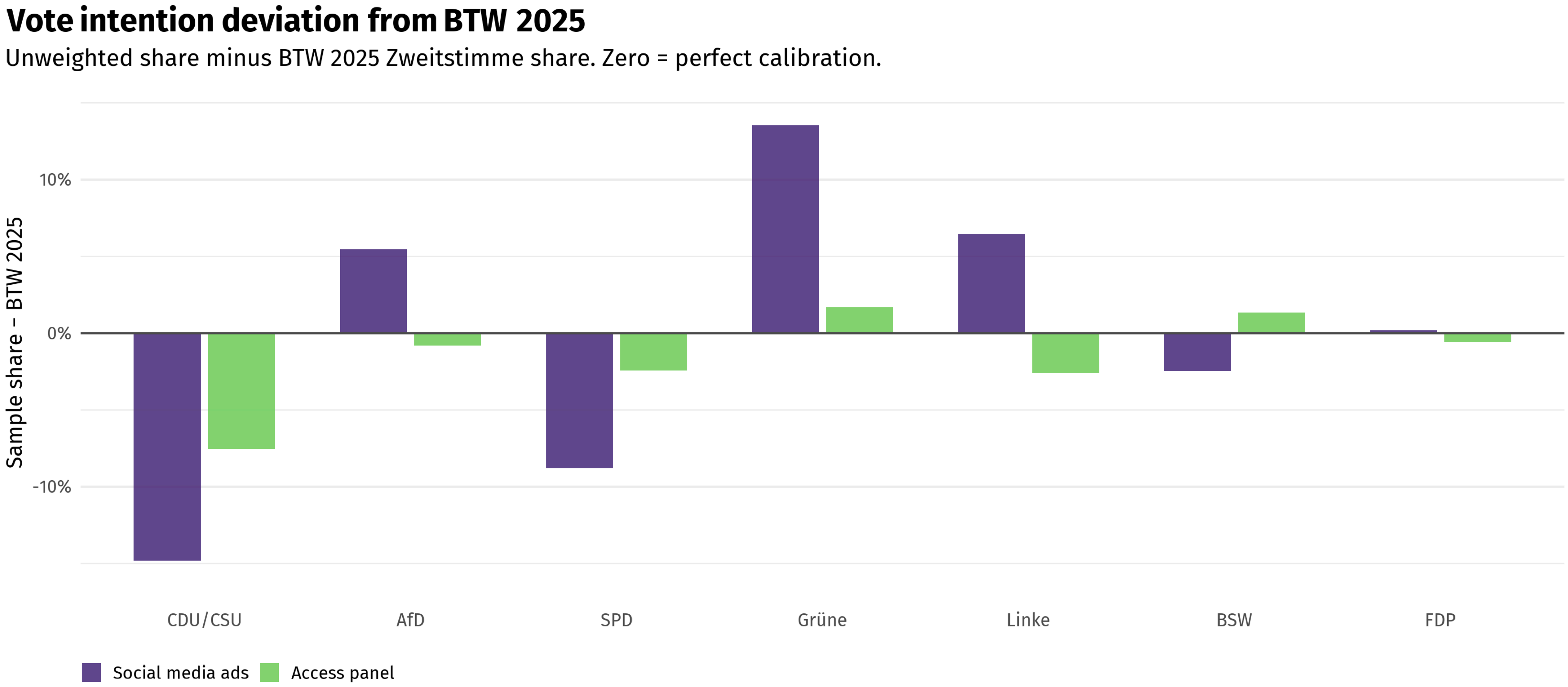
Observed share minus Mikrozensus 2023 benchmark. Zero = perfect calibration.



Sources: Zweitstimme 2025 Survey; Mikrozensus 2023 (Destatis).

Vote intention vs. official 2025 election results

Hertie School



Sources: Zweitstimme 2025 Survey; Bundeswahlleiter BTW 2025.

Measuring geographic representativeness

Measuring geographic representativeness

1. Coverage

Which units have *any* respondents? Which have *enough*?

- **Coverage rate C(k)**: share of units with ≥ k respondents

2. Proportionality, geographic concentration

Are respondent shares proportional to population shares?

- Coverage ratio: $r_i = \frac{n_i/n}{P_i/P}$; ideal: $log(r_i) = 0$ for all i
- Summarized by **Lorenz curves** and **Gini coefficient**

3. Spatial structure of imbalance

Is imbalance random noise or systematically clustered?

- **Log-ratio maps** show *where* imbalance occurs

The Modifiable Areal Unit Problem (MAUP) in geographic representativeness

Geographic representativeness is **scale-dependent**. As the geographic unit shrinks, sampling noise overwhelms signal.

Also, geographic **representativeness at a higher level might mask imbalances at a lower level**. For instance, all districts can be covered proportionally, but urban areas might still be overrepresented.

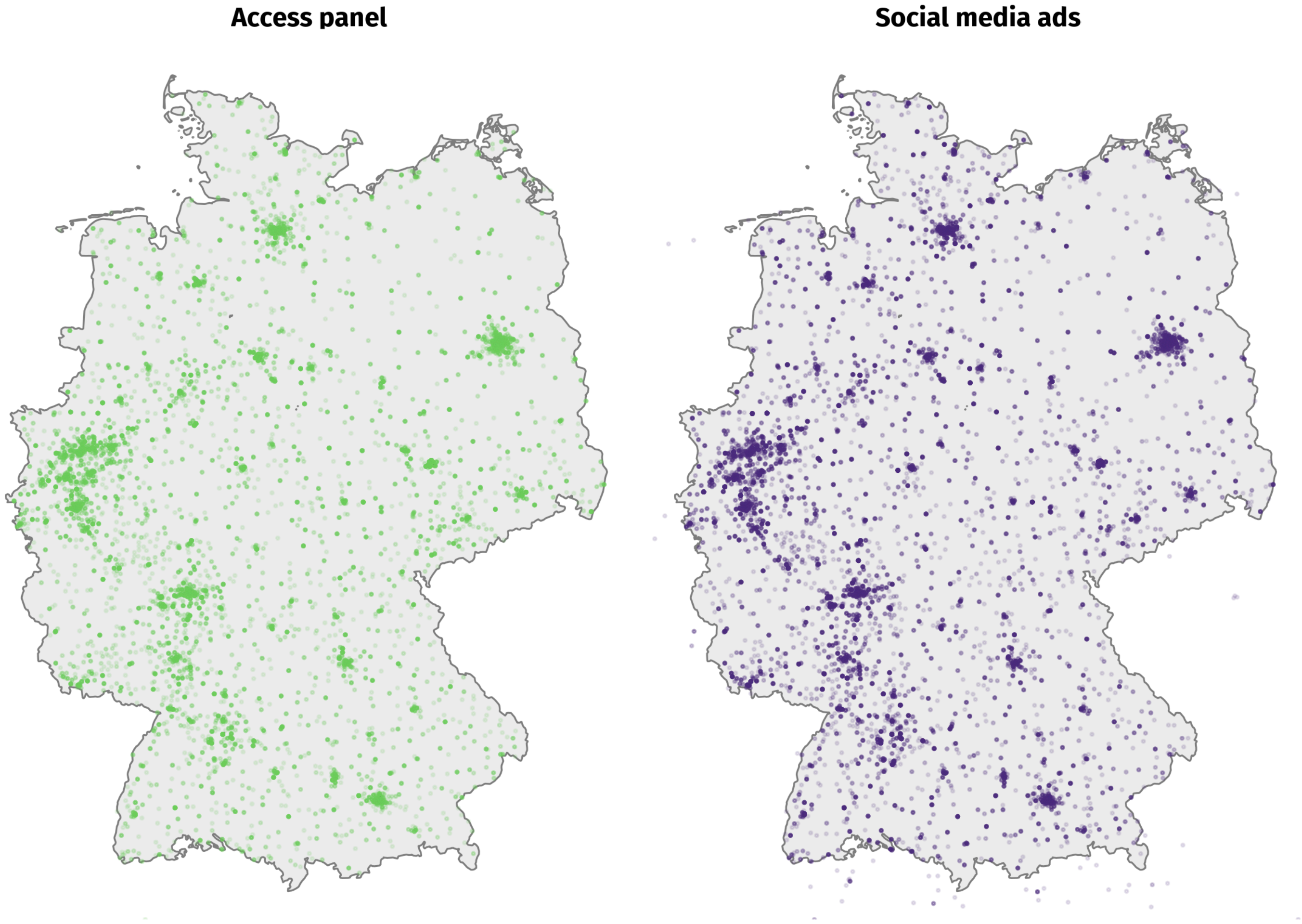
Geographic level	Units	Avg. n/unit	Key challenge
Bundesland	16	~1,500	Proportionality
Wahlkreis	299	~75	Coverage + spatial structure
Municipality	~11,000	~2	Noise dominates

IP-based respondent locations

Hertie School

Respondent locations based on IP address

Each point = one respondent. Points outside Germany border indicate non-German IP geolocation.

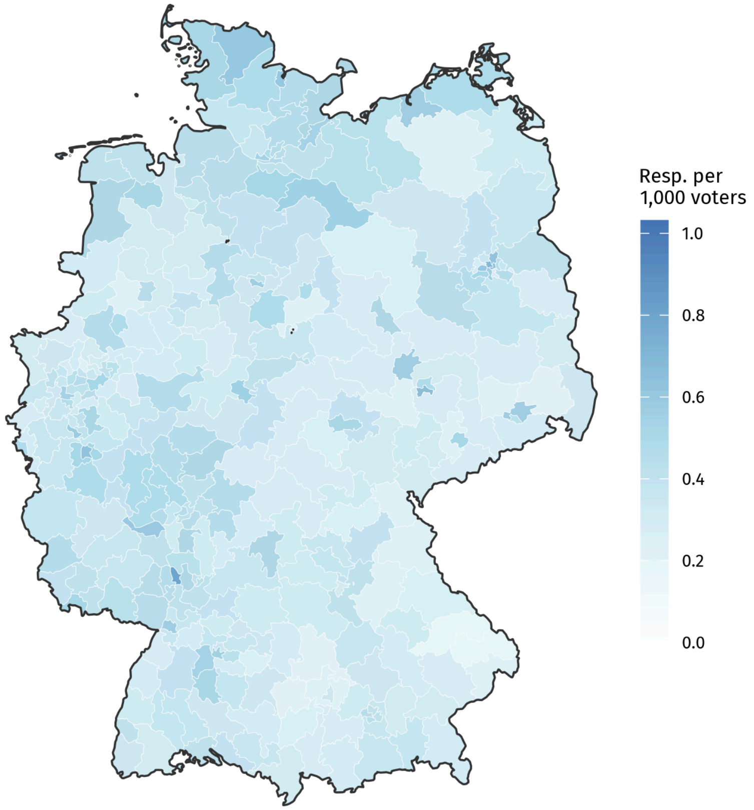


Geographic coverage maps: self-reported Wahlkreis

Geographic coverage by mode - self-reported district

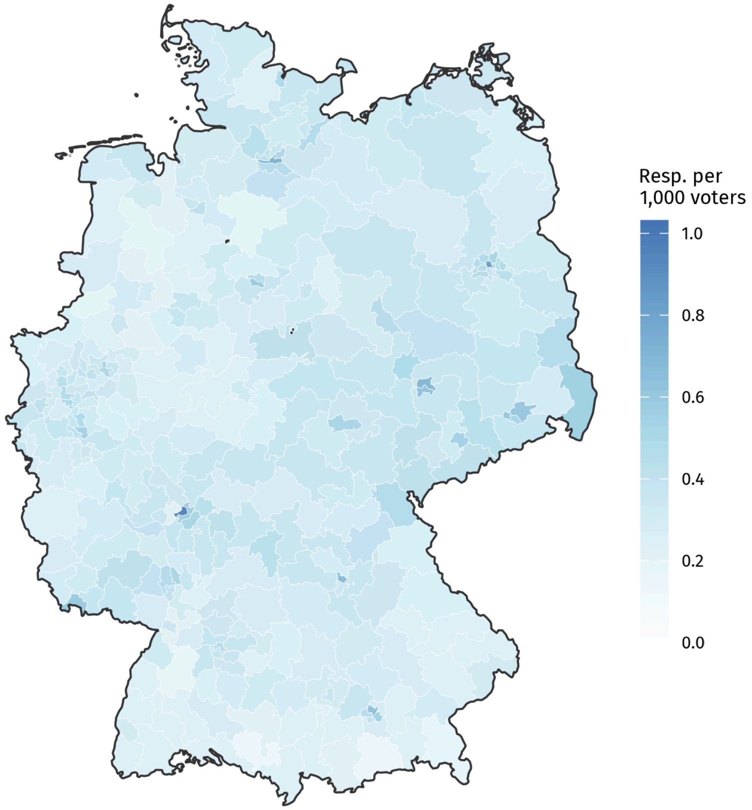
Social media ads

Respondents per 1,000 eligible voters



Access panel

Respondents per 1,000 eligible voters



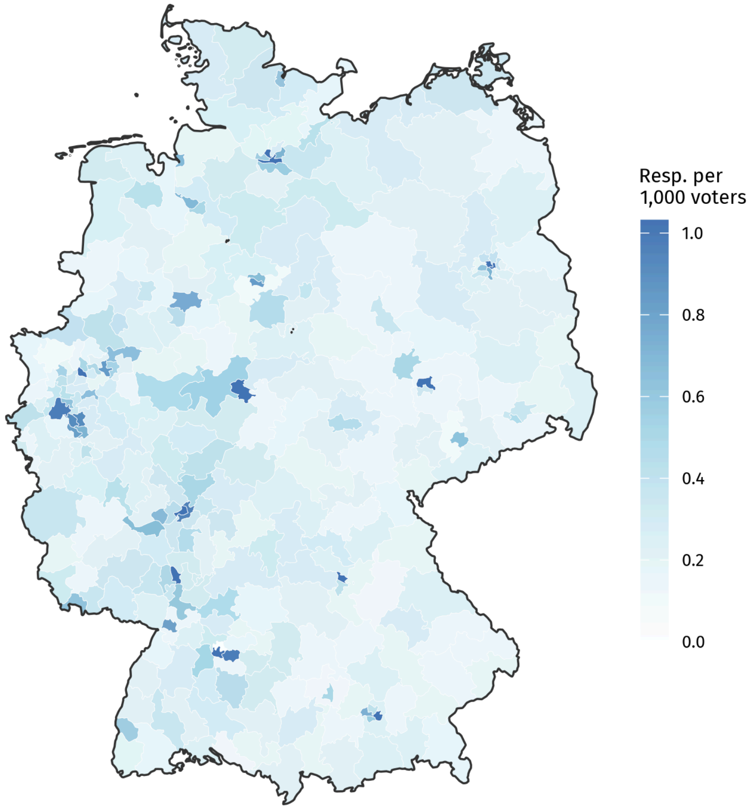
Sources: Zweitstimme 2025 Survey; Bundeswahlleiter 2025.

Geographic coverage maps: IP-based Wahlkreis

Geographic coverage by mode - IP-based district

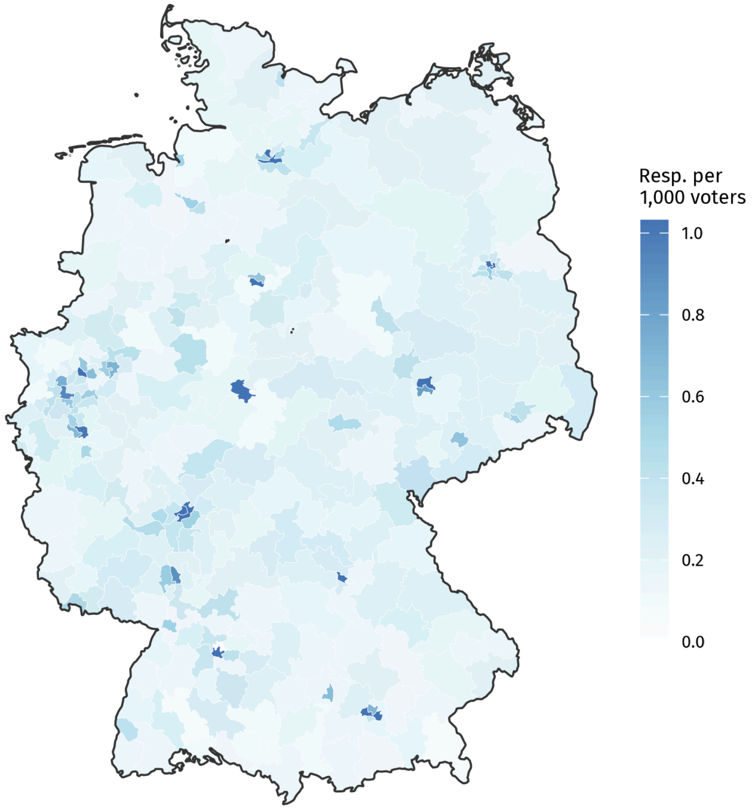
Social media ads

Respondents per 1,000 eligible voters (IP-based Wahlkreis)



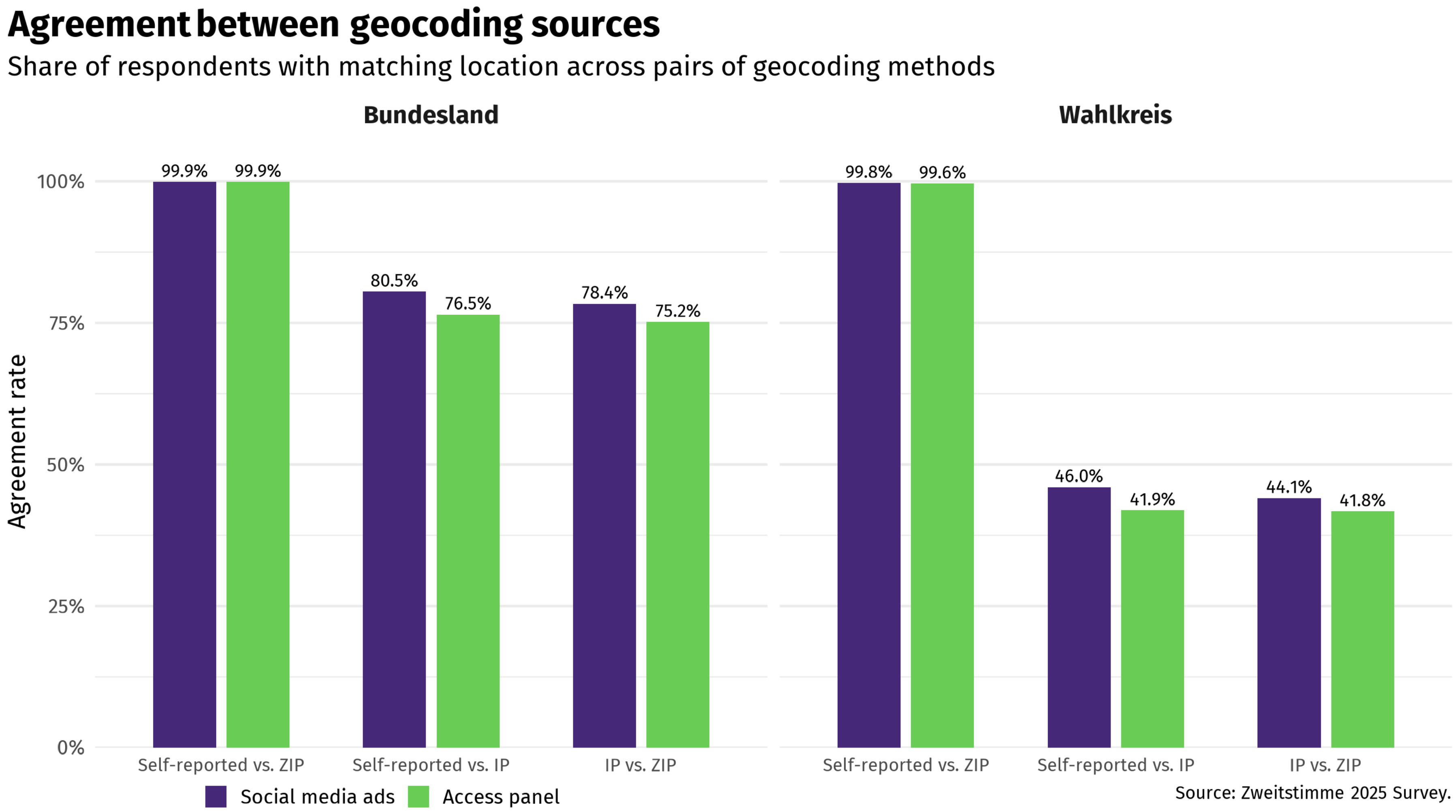
Access panel

Respondents per 1,000 eligible voters (IP-based Wahlkreis)



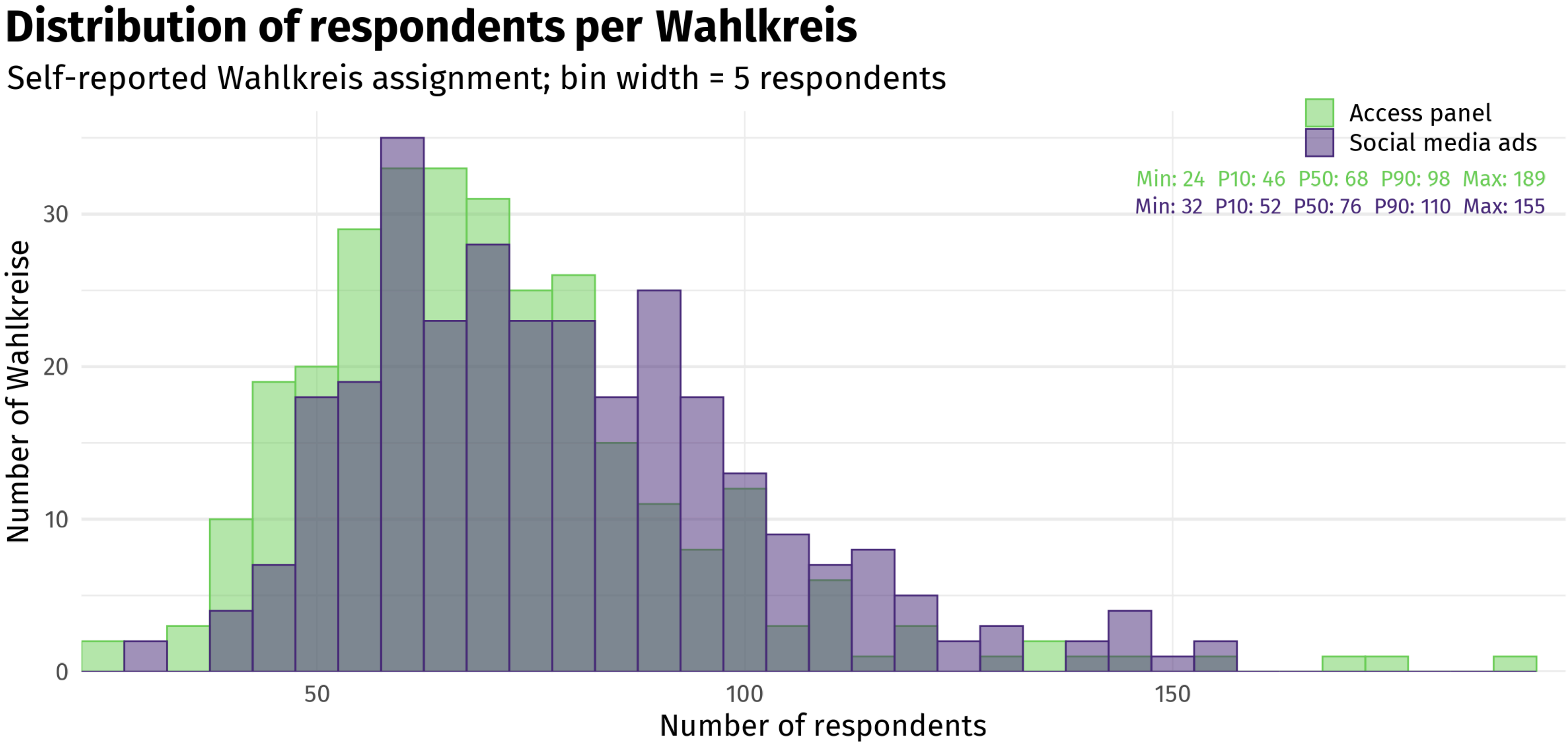
Sources: Zweitstimme 2025 Survey; Bundeswahlleiter 2025.

Agreement between geocoding sources



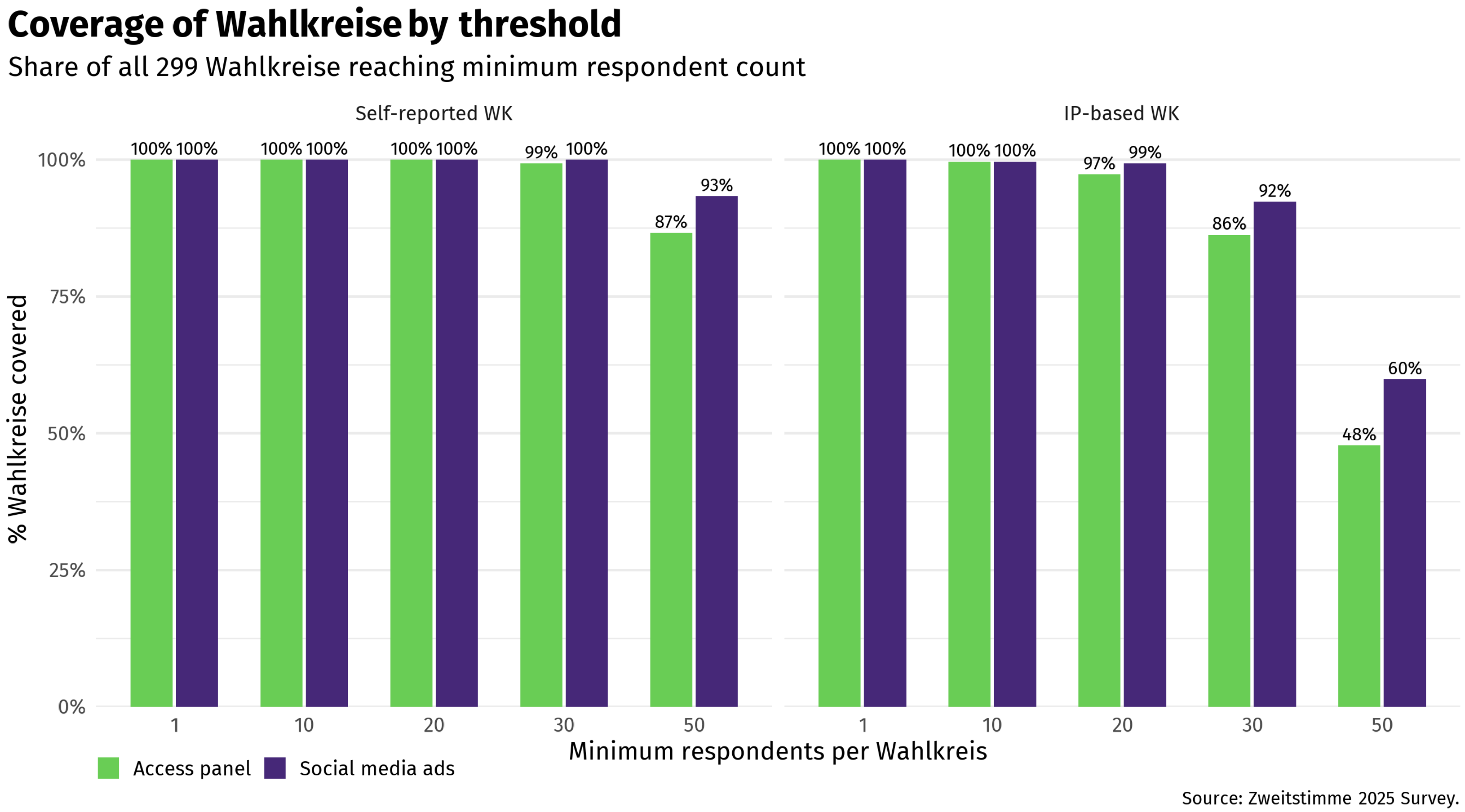
Respondents per Wahlkreis: distributional view

Hertie School

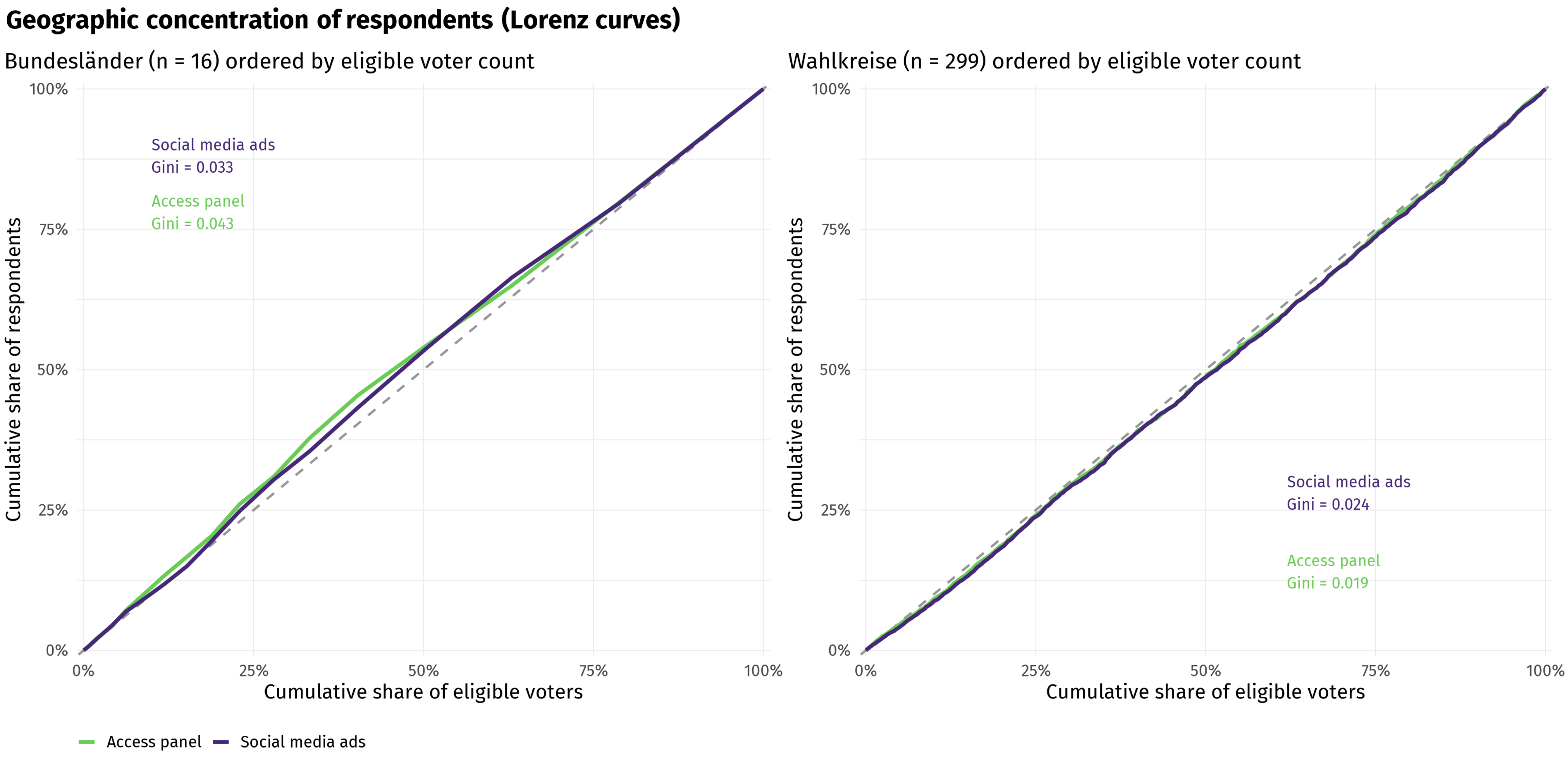


Source: Zweitstimme 2025 Survey.

Coverage rates: reaching thresholds across all 299 Wahlkreise



Geographic concentration: Lorenz curves



Source: Zweitstimme 2025 Survey; Bundeswahlleiter 2025. Diagonal = perfect proportionality.

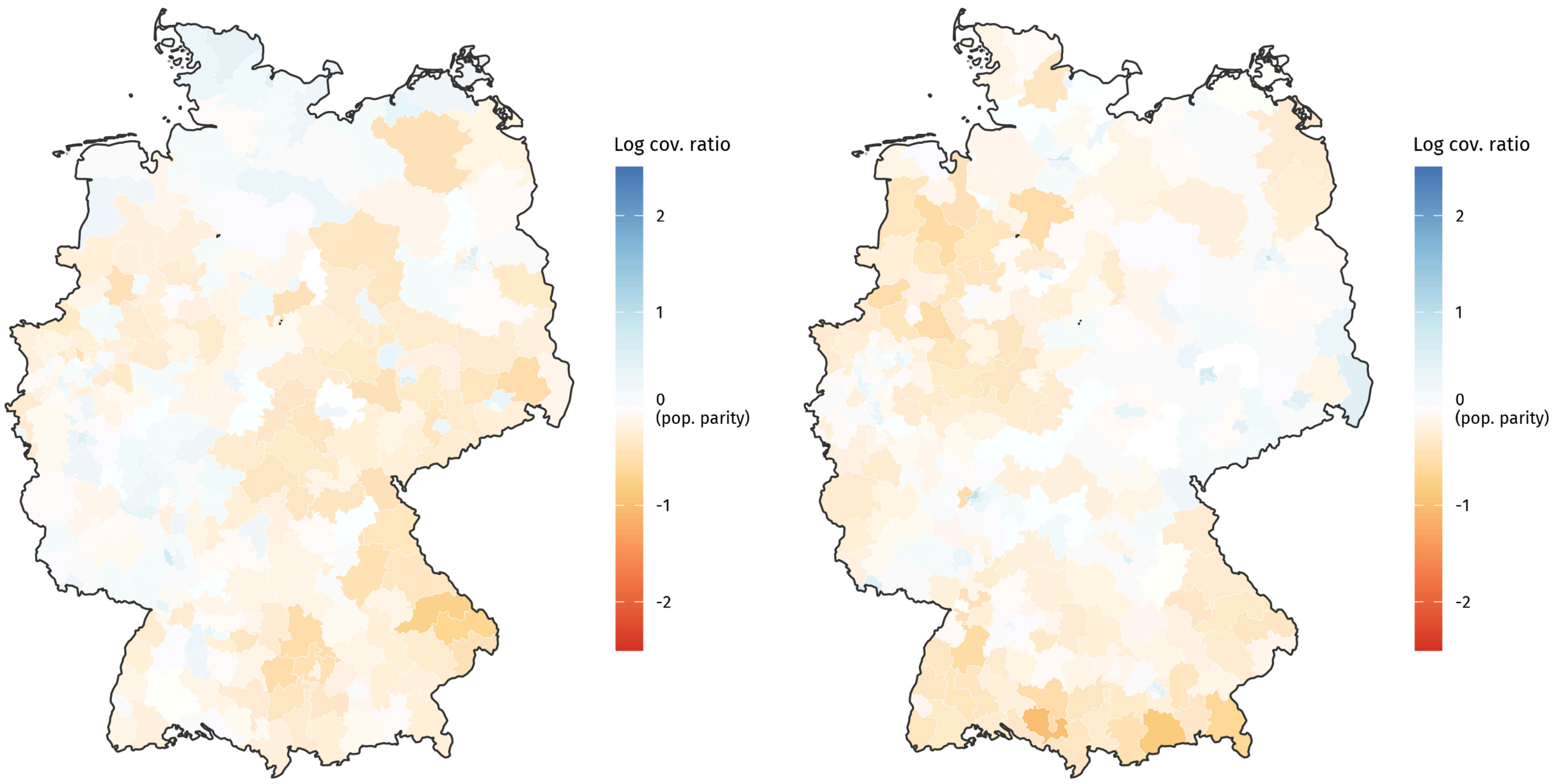
Coverage ratio map: over- and under-representation

Hertie School

Geographic over- and under-representation by mode

Log coverage ratio: positive = over-represented, negative = under-represented

Social media ads | Access panel



Sources: Zweitstimme 2025 Survey; Bundeswahlleiter 2025.

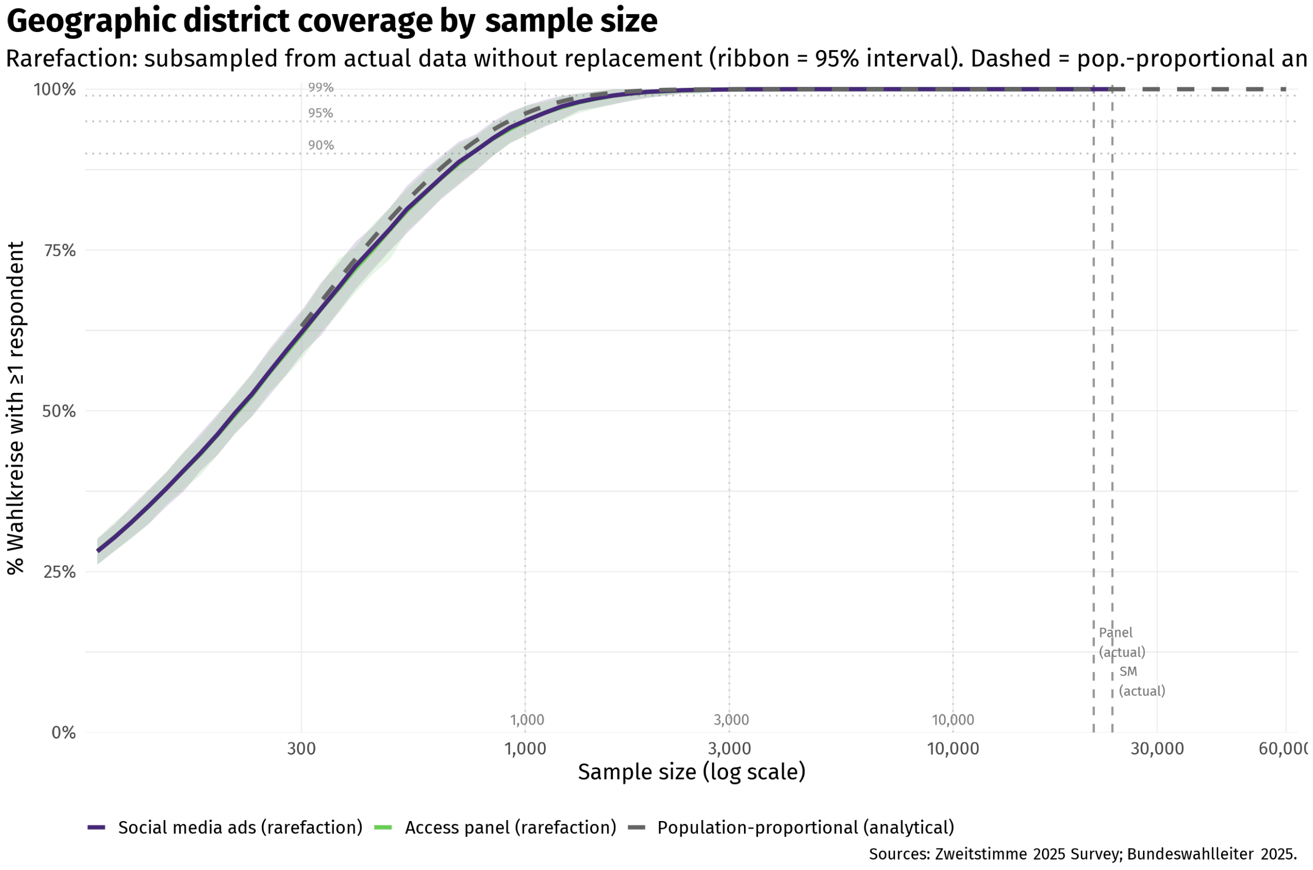
How does coverage scale?

Hertie School

Under the **population-proportional** mechanism, ≥90% WK coverage is reached at roughly **n = 2,000**; ≥99% requires ~10,000.

Under **SM-like** and **panel-like** distributions the curve is similar, because both modes approximate proportional allocation after location targeting.

Implication: For budget-constrained studies (n = 1,000–3,000), ≥95% WK coverage is achievable with social media ads in this context.



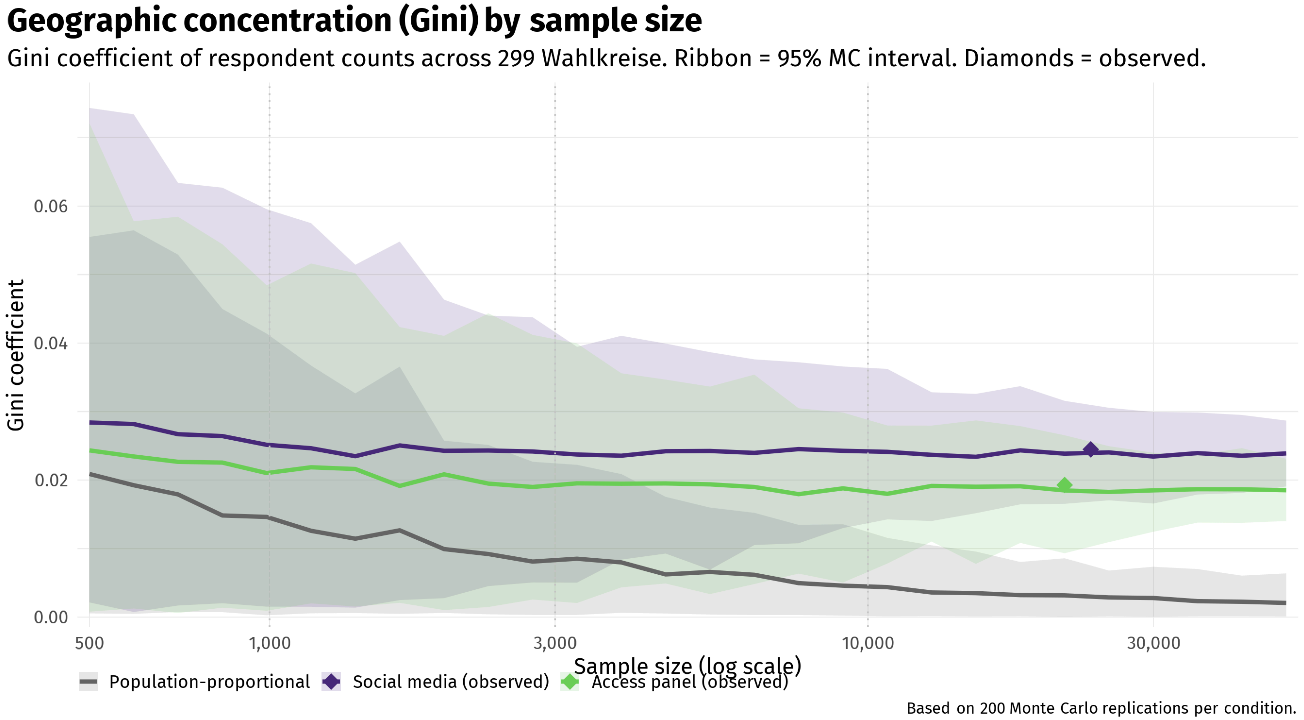
How does concentration scale?

Hertie School

The **Gini coefficient** measures geographic concentration relative to the eligible voter distribution. Zero = perfectly proportional; higher values = more concentrated.

For both modes, Gini stays **consistently below 0.03** across the full range of sample sizes — and the reference diamonds (actual n) fall on the simulated curves.

Implication: Geographic concentration is a structural feature of each recruitment mode and location targeting strategy, not a sample-size artifact.

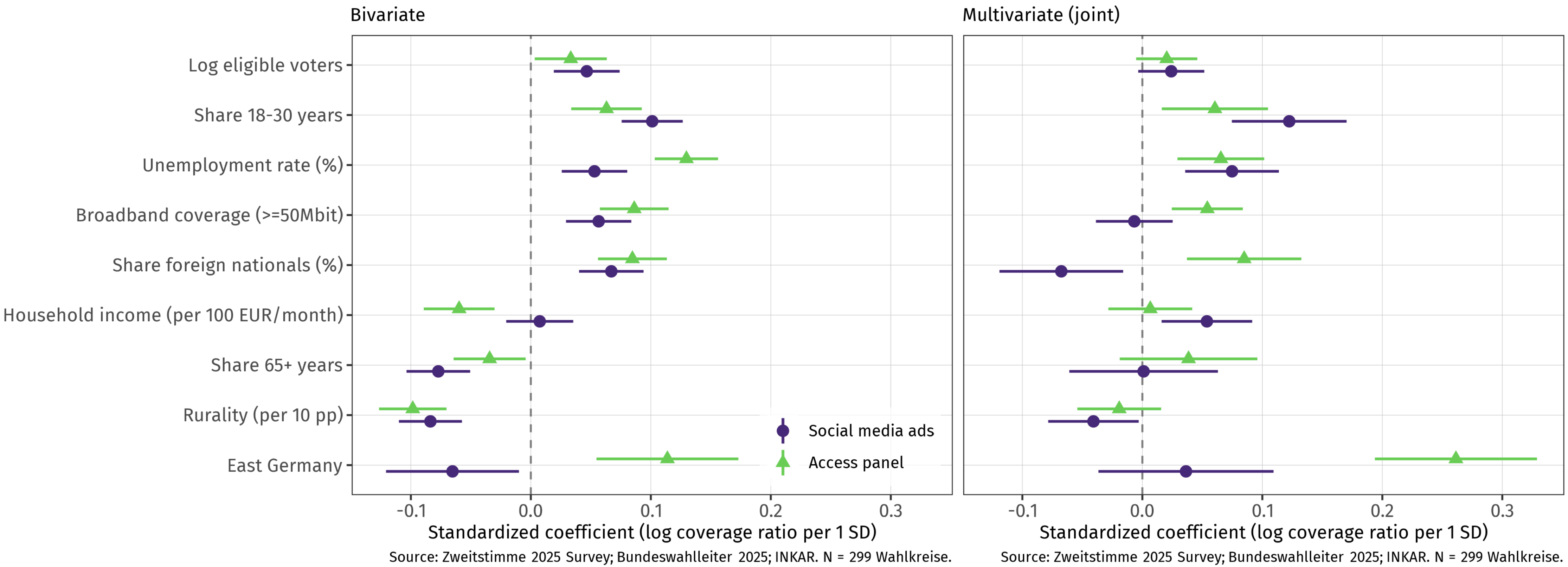


What predicts geographic coverage?

What predicts geographic over- and under-coverage?

Hertie School

Predictors of geographic over/under-representation (standardized)



OLS with z-scored predictors; binary East Germany rescaled by 2×SD (Gelman 2008). Outcome = log coverage ratio (natural scale). 95% CIs. Source: Zweitstimme 2025 Survey; Bundeswahlleiter 2025; INKAR. N = 299 Wahlkreise.

Outcome: log coverage ratio = log(respondent share / eligible voter share). Zero = proportional representation. **Predictors z-scored** (binary East Germany divided by 2×SD); coefficients reflect change in log coverage ratio per 1 SD increase — e.g., β = −0.10 → 10% lower coverage ratio; β = +0.20 → 22% higher.

When does geographic coverage affect estimates?

When does geographic coverage affect estimates?

Initial thoughts

Geographic imbalance translates into **substantive bias** when:

1. **The outcome varies geographically** — if all areas think the same, it doesn't matter who is covered
2. **Imbalance is systematic** — random geographic noise averages out; systematic under-coverage (e.g., East Germany) doesn't
3. **Demographic post-stratification can't correct it and relevant geographic markers are unavailable** — dem weighting doesn't automatically equalize geographically distinct sub-groups with otherwise identical demographics

Analytical approaches with existing data

Approach 1: MRP validation

- Use both samples to fit multilevel models with WK-level predictors
- Assess prediction accuracy for WK-level outcomes (2025 election results available for validation)

Approach 2: Subgroup analyses by geography

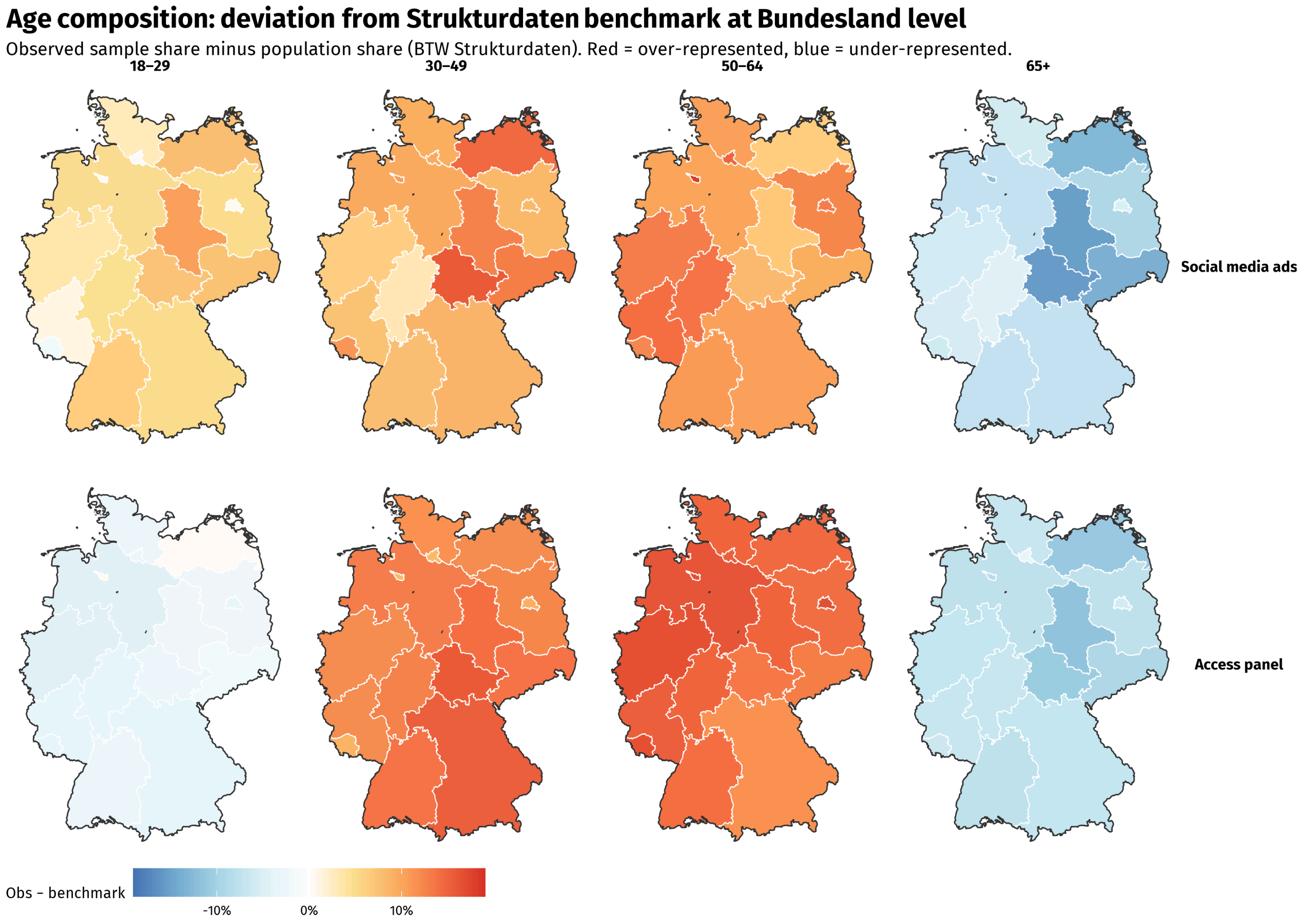
- Do treatment effects in embedded experiments (e.g., forecast exposure experiment) vary by high/low coverage districts?

Approach 3: Evaluation of weighting strategies

- Compare estimates with and without geographic raking at different levels to explore downstream effects on WK-level and individual-level estimates

State-level age composition: deviation from population benchmarks

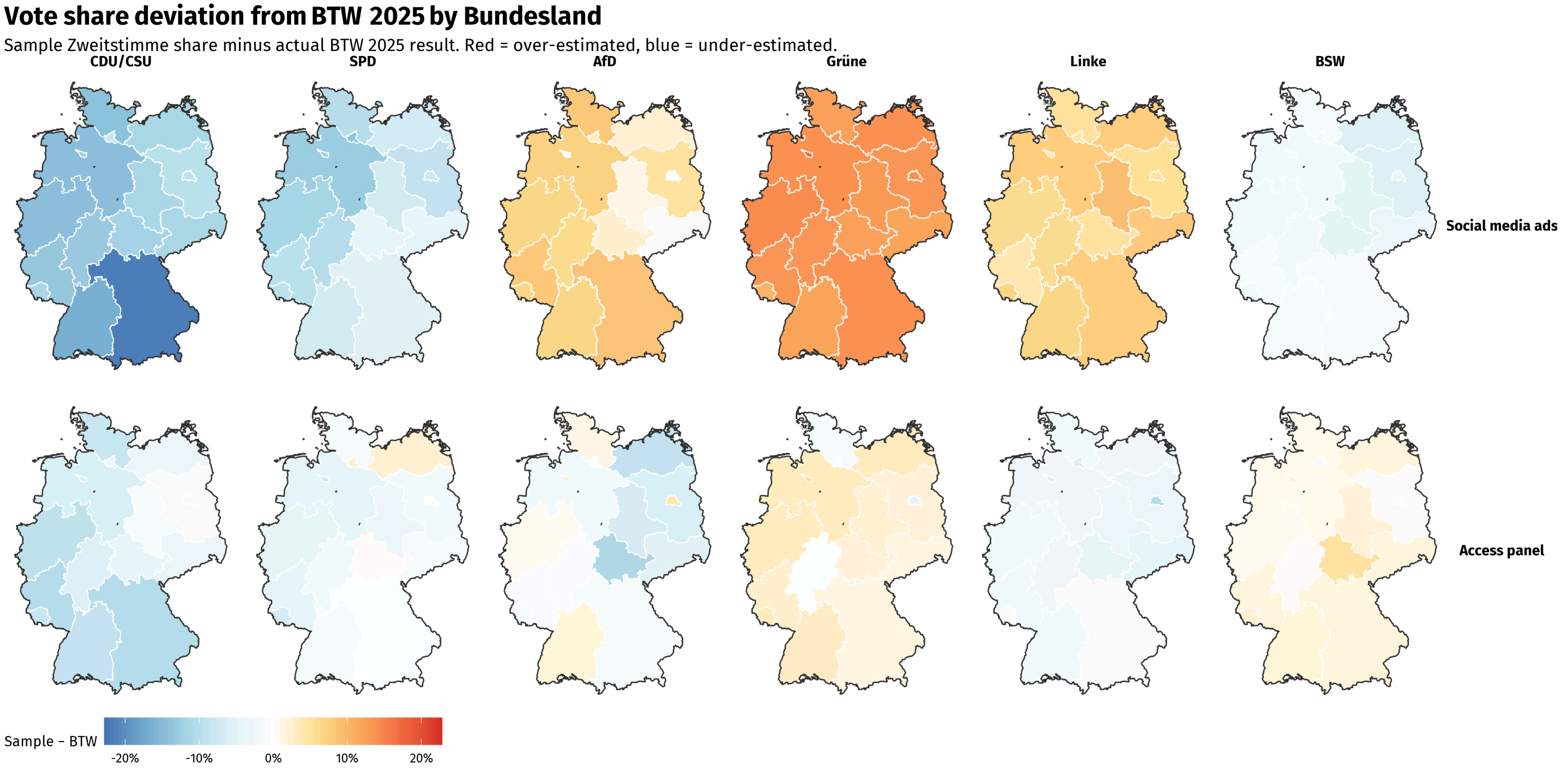
Hertie School



Sources: Zweitstimme 2025 Survey; Bundeswahlleiter BTW 2025 Strukturdaten.

State-level vote shares: deviation from official 2025 results

Hertie School



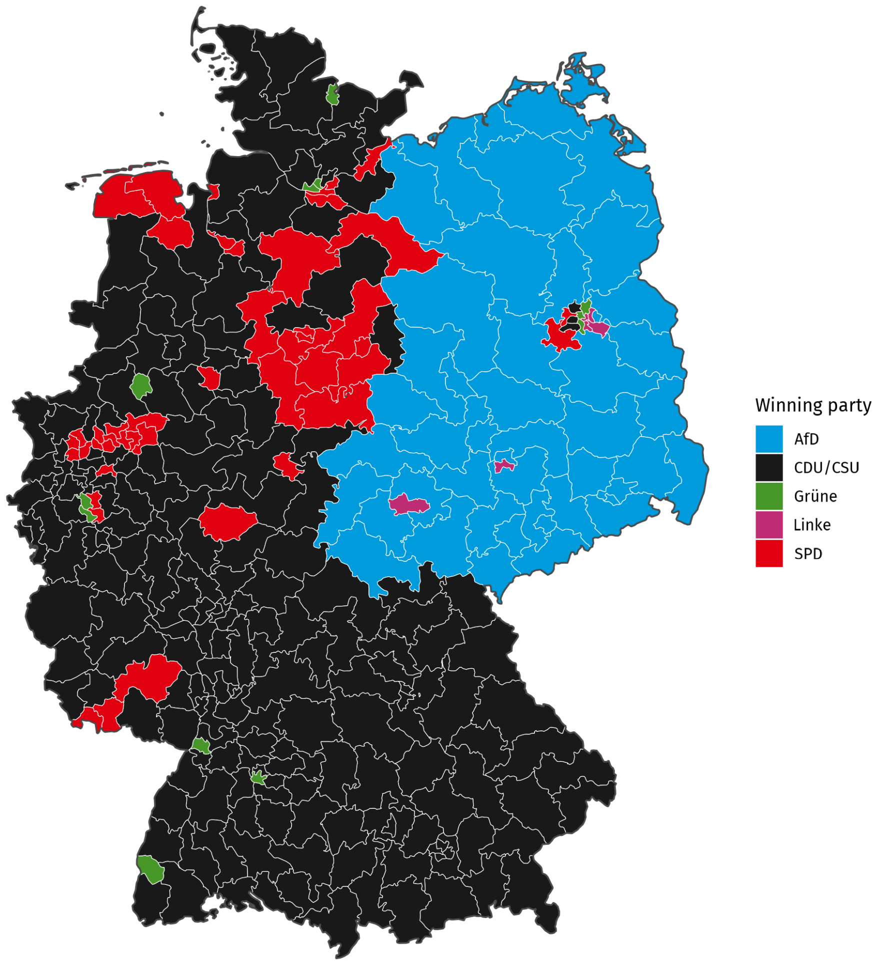
Sources: Zweitstimme 2025 Survey; Bundeswahlleiter BTW 2025.

District-level results

Hertie School

Erststimme district winners - BTW 2025

Winning party in each of the 299 Wahlkreise



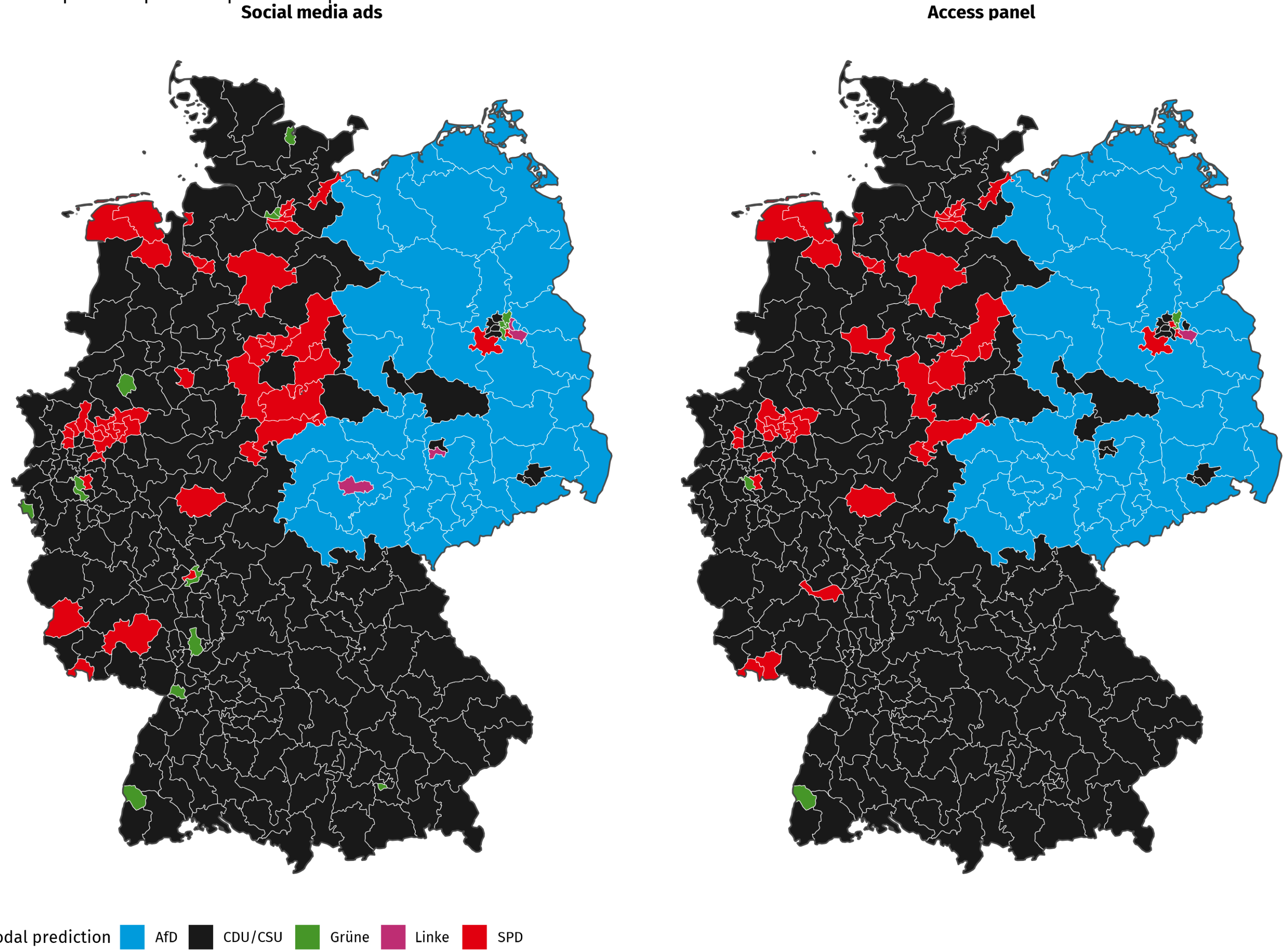
Source: Bundeswahlleiter 2025.

District-level citizen forecasts

Hertie School

Modal predicted Erststimme winner by recruitment mode

Most frequent respondent prediction per district.



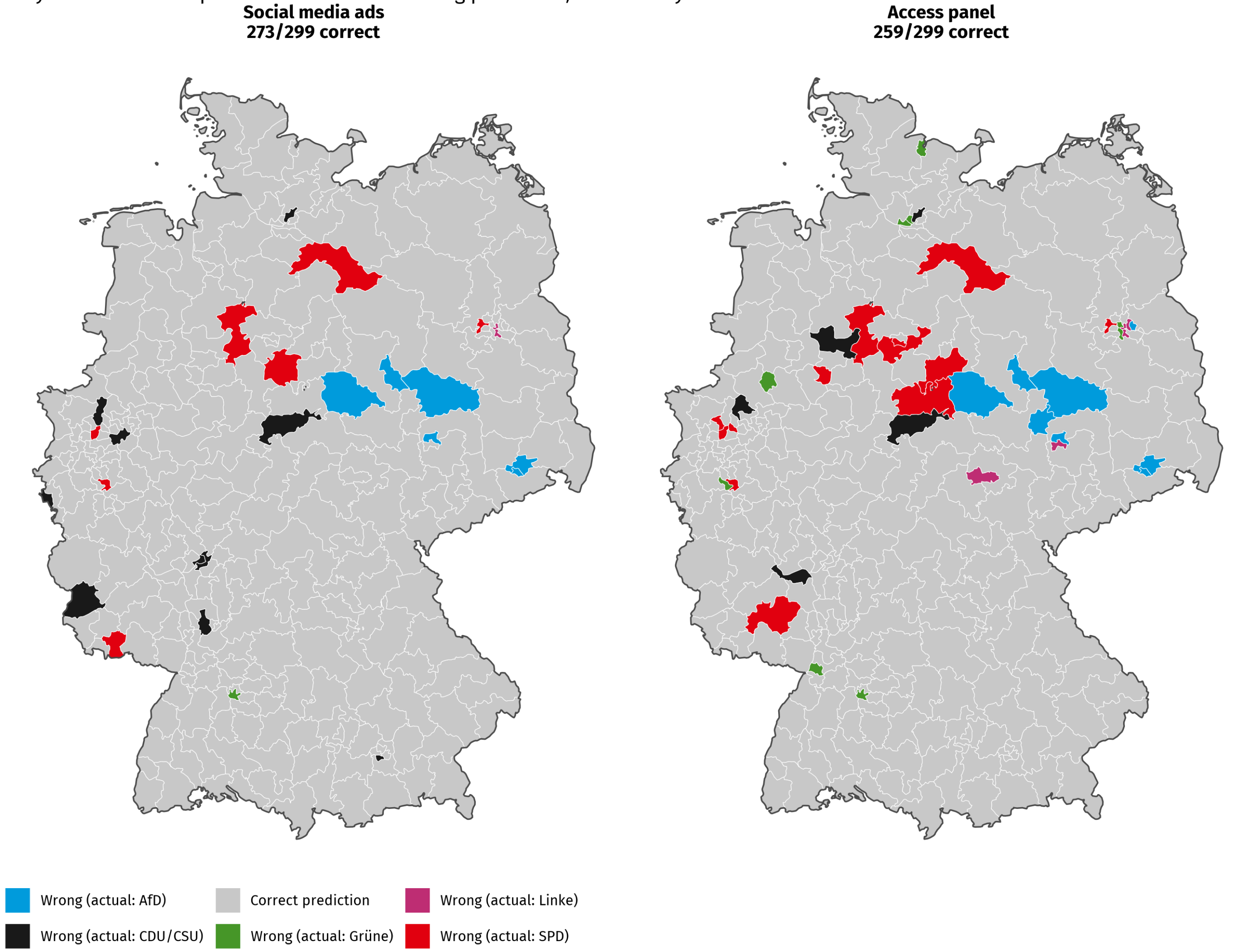
Source: Zweitstimme 2025 Survey.

District-level citizen forecasts

Hertie School

Erststimme winner mispredictions by recruitment mode

Grey = correct modal prediction. Coloured = wrong prediction, coloured by actual winner.



Source: Zweitstimme 2025 Survey; Bundeswahlleiter 2025.

Take-aways

Summary and take-aways

On geographic representativeness

- Geographic representativeness is a **distinct, underappreciated dimension** of survey quality — not reducible to demographic balance
- **Measuring** geographic coverage requires multiple methods: self-report, IP geolocation, and ZIP crosswalk each have strengths and weaknesses

On social media recruitment

- Facebook/Instagram ads achieve **universal Wahlkreis coverage** via state-level targeting at manageable cost
- **No clear East/West coverage gap** (in comparison to the access panel)
- **IP geolocation** provides useful Bundesland-level accuracy but might be too noisy for district-level analyses

On what predicts geographic coverage for social media surveys

- Positive: Younger, more urban, higher income, unemployment rate
- Negative: Rurality, share of foreign-born residents
- Broadband coverage? Nope
- Social media usage? Likely; not measured at district level

Bottom line

If geographic representativeness matters for your research question, **social media recruitment with state-level targeting** is a cost-effective complement or alternative to traditional access panels.

Next steps and open questions

This project

- Complete **MRP validation** of district-level vote intention estimates against final 2025 election results
- Downstream **experimental subgroup analyses** (forecast exposure treatment × geographic coverage)
- **Toolbox paper**: practical guide to geographic targeting on Meta for social scientists

Methodological extensions

- Can **within-state ZIP-code targeting** on Meta improve sub-state geographic uniformity?
- How do **post-stratification schemes that include geographic predictors** (not just demographics) change estimate quality?

Thank you!

Questions, comments, feedback welcome.

munzert@hertie-school.org

zweitstimme.org

Appendix

IP-based geolocation: how it works and how much to trust it

Hertie School

The mechanics

1. Survey platform records the **device IP address** at time of response
2. IP → approximate geographic coordinates via a **MaxMind- or ip-api-style database** (maps IP blocks to lat/lon centroids)
3. Coordinates → administrative unit via **spatial join** (point-in-polygon on shapefile)

Sources of error

- **ISP routing**: IP blocks registered to ISP headquarters, not user location — common for mobile and broadband
- **VPNs and proxies**: user location is masked; IP resolves to VPN exit node
- **IPv6 and CG-NAT**: shared addresses complicate assignment
- **Dynamic IPs**: reassigned addresses may carry stale location records
- **Spatial resolution**: ~city-level accuracy (~25–50 km median error in Europe); sub-city assignment adds further noise

What the literature says

Accuracy benchmarks (Europe)

- Country level: >99% correct (Poese et al. 2011)
- Region/NUTS-2: ~85–90% (varies by database and country)
- City level: ~60–80% within 25 km
- Sub-city / district: often <50% — not reliable as primary measure

Mobile vs. fixed line

- Mobile IPs are harder to localize — GPS location ≠ IP location
- Carrier routing concentrates mobile IPs in a few cities → systematic urban bias

Practical implications

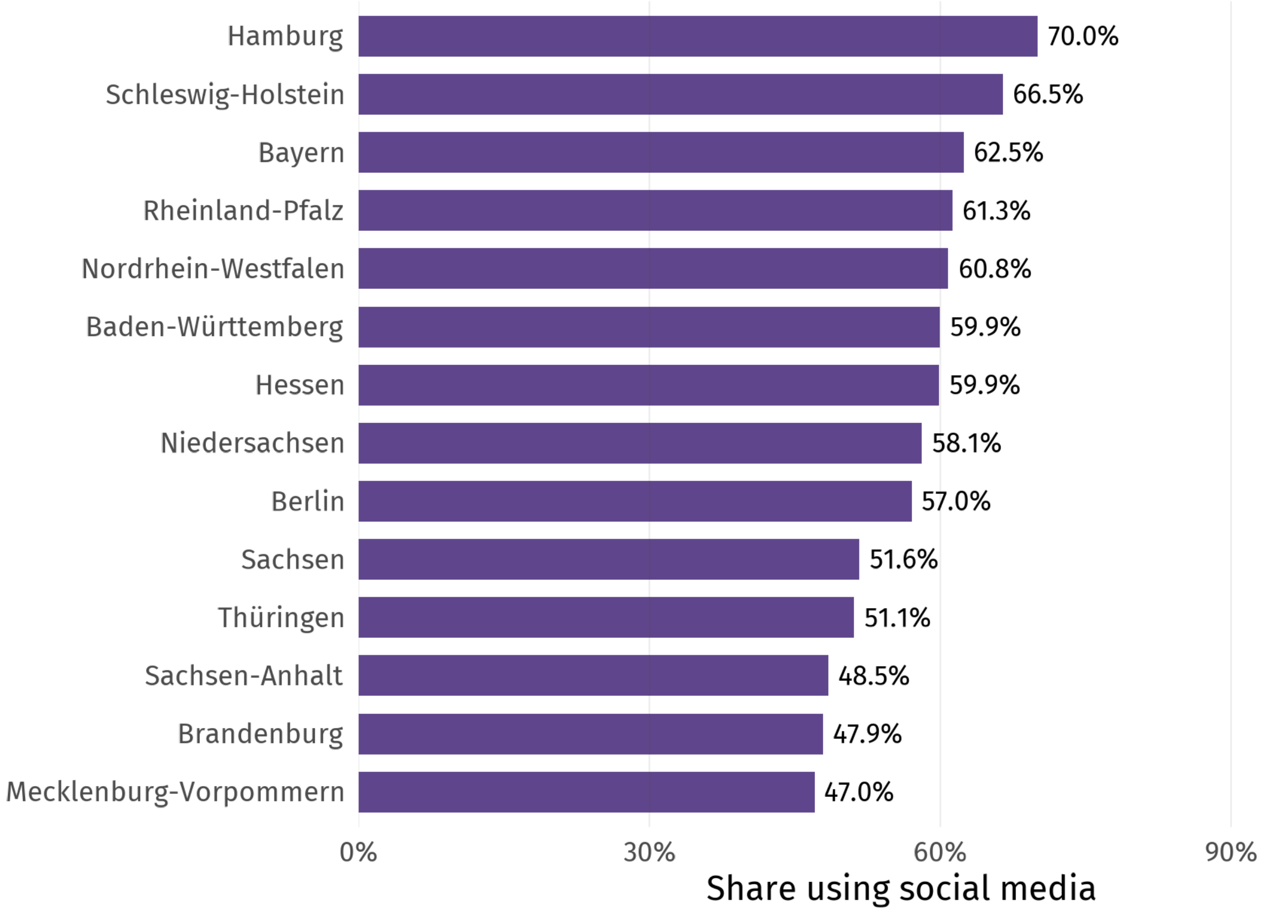
IP geolocation is always available and **reliable at the Bundesland level** (~80% agreement with self-report in our data). At the **Wahlkreis level it is too noisy** to use as a primary measure.

Social media usage by Bundesland

Hertie School

Social media usage by Bundesland

Share of population aged 16–74 who used social networks in the past 3 months.
14 of 16 Bundesländer reportable.



Source: Destatis Genesis 12231-0101, IKT-Nutzung in privaten Haushalten 2025.

Does reweighting fix geographic imbalance?

Hertie School

Three weighting strategies

Unweighted — raw respondent distribution

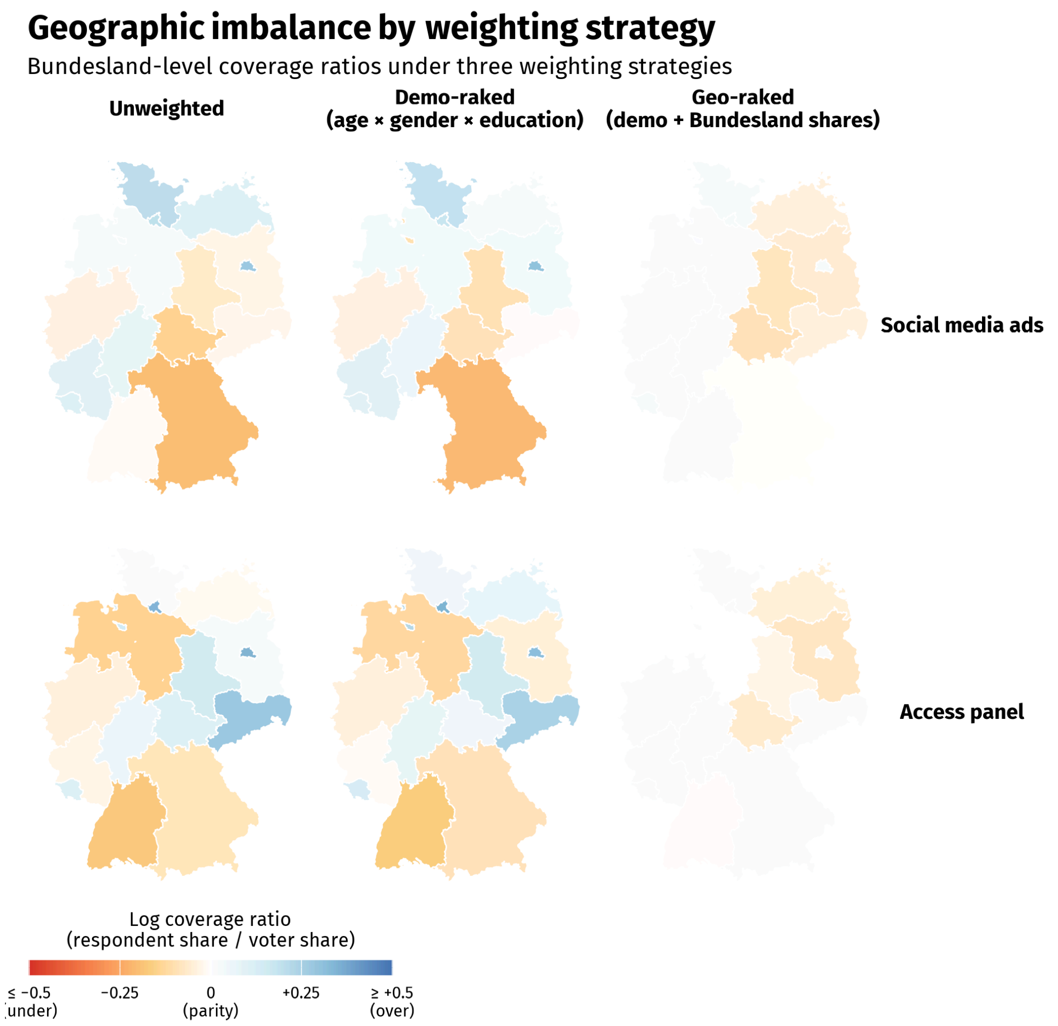
Demo-raked — iterative proportional fitting within each of the 16 Bundesländer on **age × gender × education**; targets: Mikrozensus 2023. DEFF ≈ 2.

Geo-raked — demo-raking + second rake pass at national level adjusting **Bundesland shares** to eligible voter distribution (Bundeswahlleiter 2025).

Key finding

Demographic raking leaves geographic imbalance **nearly intact** — the first two columns look essentially identical.

Adding Bundesland shares as an explicit raking margin **eliminates the geographic bias** (last column).

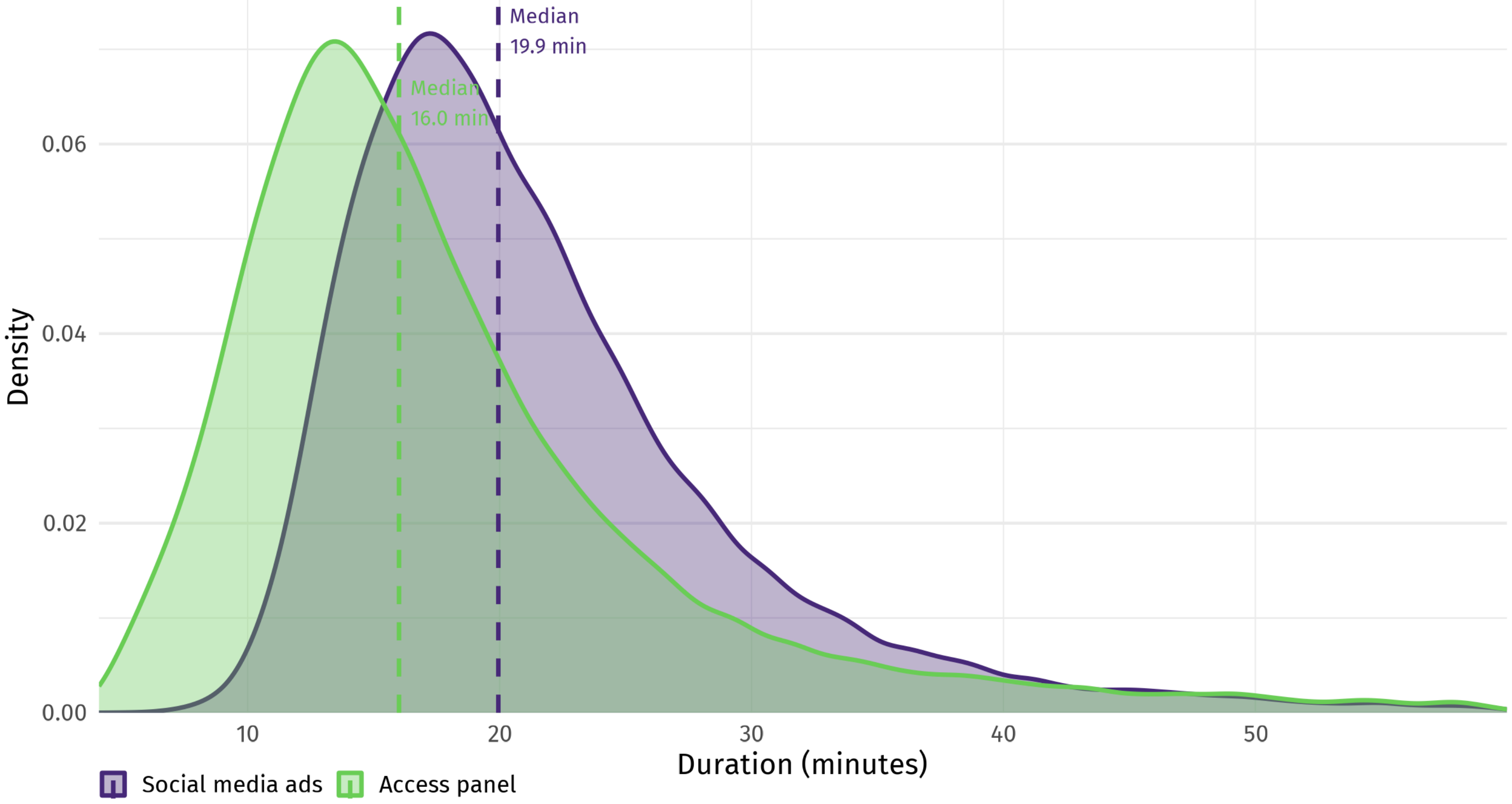


Sources: Zweitstimme 2025 Survey; Bundeswahlleiter 2025.

Survey duration distribution

Survey duration distribution by mode

Truncated at 60 minutes; dashed lines = medians



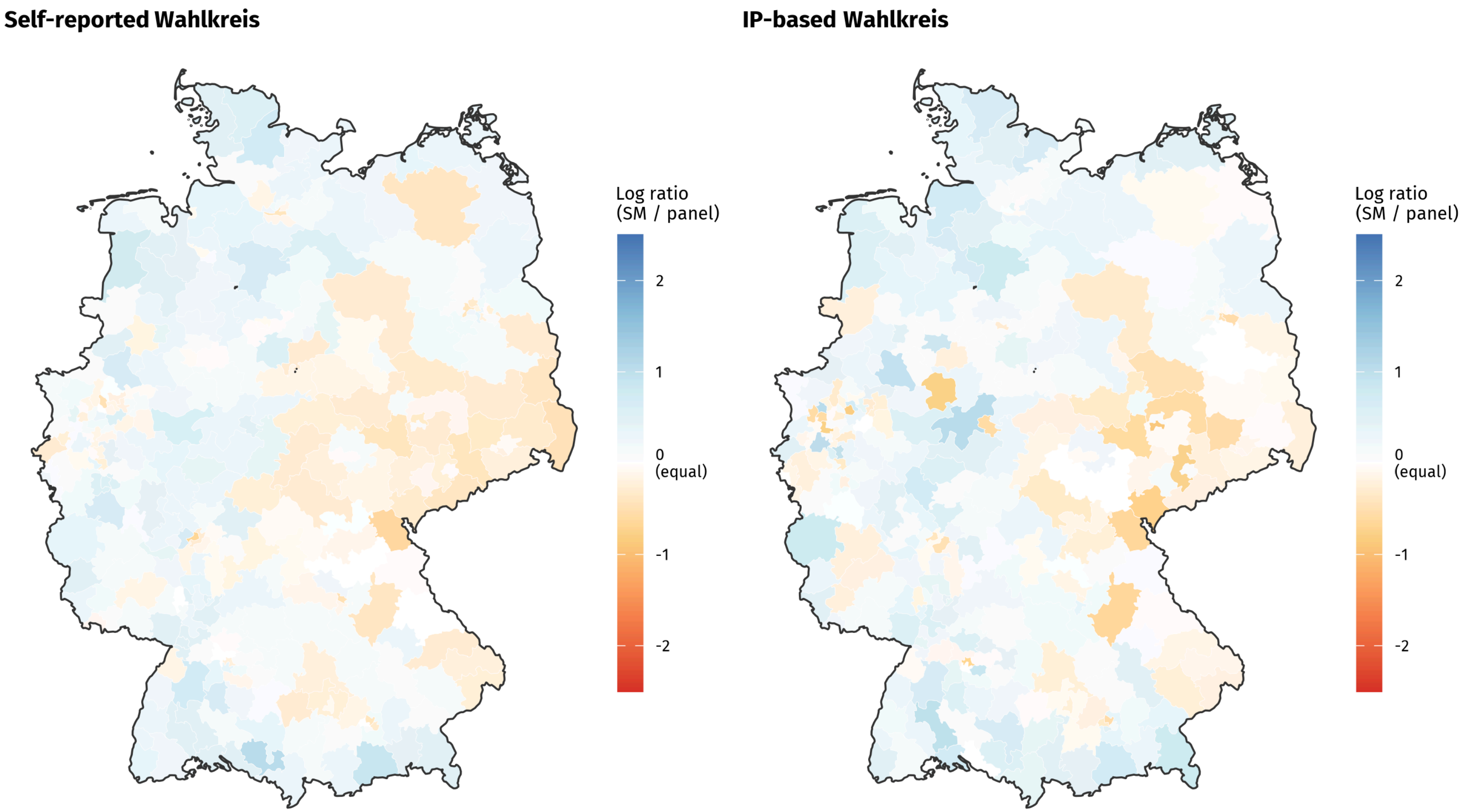
Source: Zweitstimme 2025 Survey.

Social media vs. access panel: relative geographic coverage

Hertie School

Social media vs. access panel: relative geographic coverage

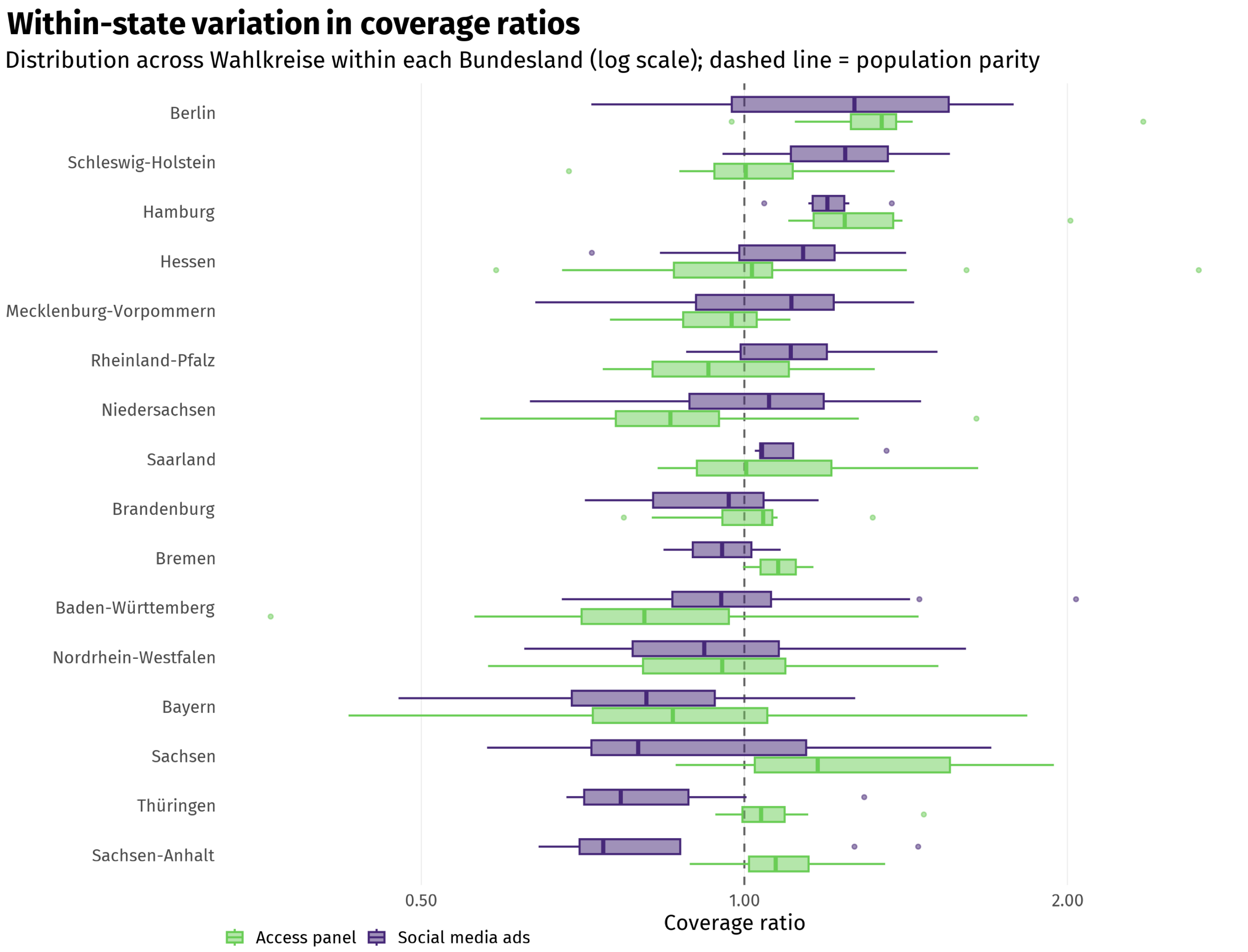
Log ratio > 0: SM over-represents district relative to panel; < 0: SM under-represents



Sources: Zweitstimme 2025 Survey; Bundeswahlleiter 2025.

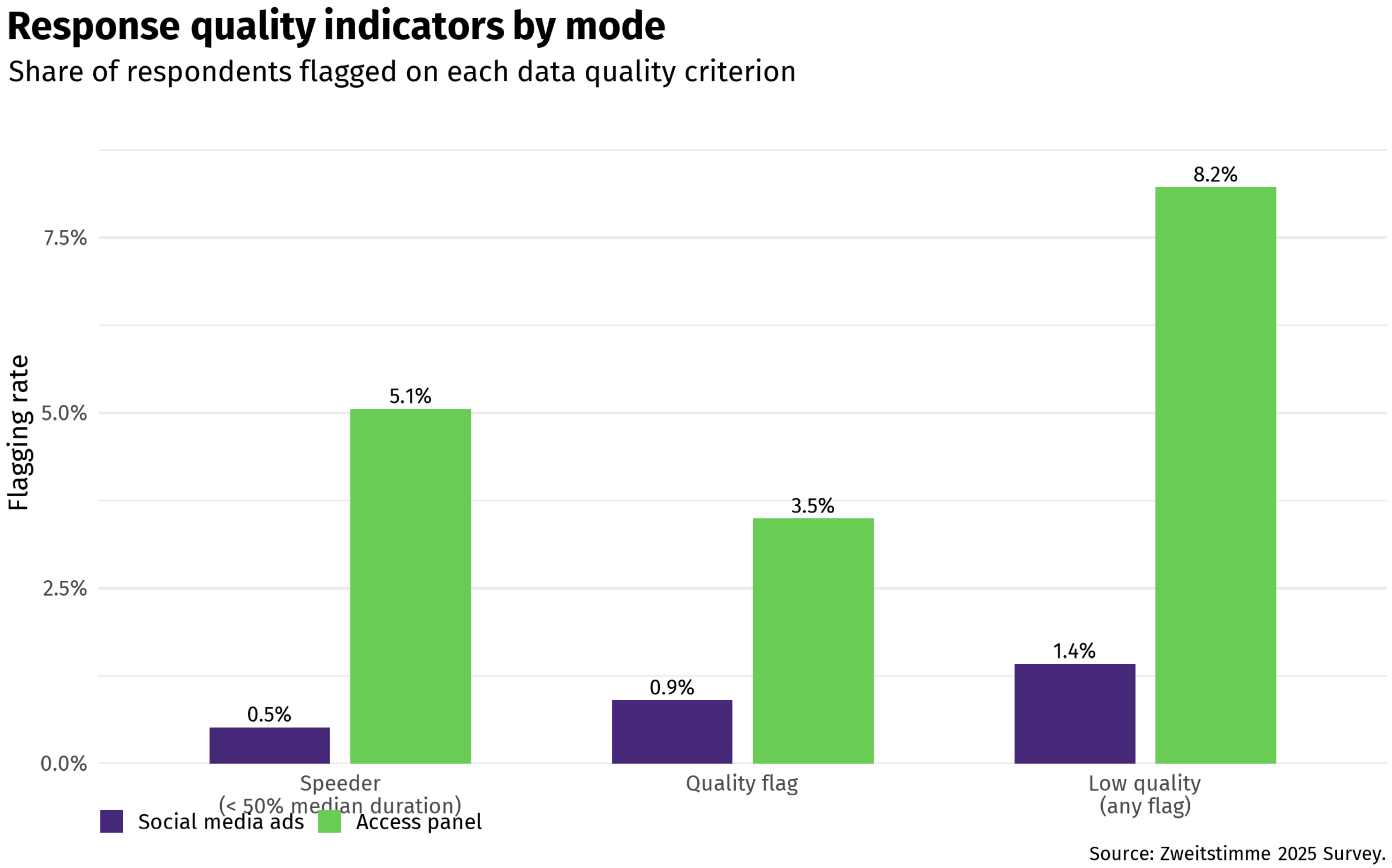
Within-state variation in coverage ratios

Hertie School



Source: Zweitstimme 2025 Survey; Bundeswahlleiter 2025.

Response quality by mode



Response quality by demographics