Administrative Failure, State Capacity, and Democratic Exclusion: Evidence from Berlin's 2021 Election Breakdown*

Marius Kröper†

October 27, 2025

Preliminary Draft — Please do not cite or distribute

[latest version]

Abstract

This paper studies the long-run effects of non-strategic administrative failures on voter participation. I exploit a natural experiment from Berlin's 2021 elections, in which hundreds of precincts experienced ballot shortages, multi-hour queues, and unlawful polling closures. Using precinct-level administrative data and a stacked event study design, I show that precincts exposed to administrative failures in the 2021 Berlin election experienced a 1.8 percentage points (2.4%) decline in turnout across three subsequent elections over the next four years. The drop is concentrated in in-person voting and only partially offset by increases in postal participation in subsequent elections. Effects are largest among young voters, welfare recipients, and residents with migration backgrounds. Survey evidence suggests two mechanisms: disrupted civic habit formation and short-term erosion of institutional trust.

Keywords: State Capacity, Voter Turnout, Voting Costs, Administrative Failure

JEL Codes: D72, H11, H70, R50

*I gratefully acknowledge the invaluable feedback from Alexander Kemnitz, Enrico Cantoni, Valentin Lindlacher, Giacomo Opocher, and the participants of the TUD Brown Bag Seminar, the RGS 2025 conference, the Potsdam PhD Workshop in Empirical Economics, and the 2025 Annual Meeting of the Verein für Socialpolitik. All remaining errors are my own.

†TUD Dresden University of Technology (TU Dresden). Helmholtzstr. 10, 01069 Dresden (Germany). marius.kroeper@tu-dresden.de.

1 Introduction

The quality of core public service delivery shapes citizens' trust in the state and their willingness to participate in collective decision-making. Elections are one such service. Most work on voter participation emphasizes formal barriers to access or strategic efforts to mobilize and demobilize particular groups. Much less attention has been paid to routine administrative competence: whether the state can execute the basic logistics of voting. Failures in implementation—ballot shortages, unlawful polling-place closures, multi-hour queues—are not written into law, but they are experienced by voters as direct evidence of the state's capacity and priorities. If these failures persist or are perceived as targeted, they may not only generate immediate disenfranchisement on election day, but also produce long-run withdrawal from democratic participation.

This paper studies the downstream effects of *non-strategic* administrative failure. I examine the consequences of severe procedural disruptions during Berlin's September 26, 2021 federal elections, when 431 precincts experienced ballot misallocations, hours-long queues, and temporary polling-place closures that violated electoral law. These disruptions were not partisan interventions, and they did not reflect legal restrictions on access. They were failures in the delivery of a core public service. Prior work shows that long wait times at polling places in the United States depress future turnout (Pettigrew, 2021). I show that similar downstream effects arise even in a high-capacity, high-trust setting where overt strategic suppression can be ruled out, and that these effects are persistent.

Using quasi-experimental variation and dynamic difference-in-differences design based on court-validated treatment assignments, I estimate the causal impact of exposure to these disruptions on subsequent electoral participation. Treated precincts exhibit a sustained decline of 1.8 percentage points (2.4%) in total turnout that persists for at least four years, with no evidence of recovery. The decline is driven by a sharp and lasting reduction in in-person voting and is only partially offset by increased postal voting. Because Berlin's election-day failures became a citywide scandal, even voters in formally "untreated" precincts were exposed to information about misadministration. This implies that the estimated turnout losses are a lower bound on the true effect of direct exposure to administrative breakdown.

My identification strategy exploits the fact that the German Federal Constitutional Court retrospectively reviewed the 2021 election and formally annulled local results in specific precincts. I define a precinct as "treated" if the Court documented severe procedural violations—including missing or incorrect ballots, legally impermissible polling-place closures, or queue times extending well beyond the statutory voting window—and ruled that these violations were serious enough to invalidate the local outcome. Crucially, precincts were not labeled as treated because of low turnout, partisan imbalance, or ex post complaints by losing candidates. Treatment status is therefore based on observed administrative breakdowns, not on realized electoral results. This court-defined classification allows comparisons across precincts within Berlin that are otherwise subject to the same electoral rules, political environment, and media coverage, but that did or did not experience a documented failure in service delivery. Baseline turnout levels and

demographic composition in treated precincts were similar to those in untreated precincts, and pre-trends in turnout are parallel.

I interpret these results as evidence that elections function as recurring tests of state capacity. The contribution of this paper is to show that *non-strategic* bureaucratic failures in the administration of voting can generate persistent disengagement, concentrated among already underrepresented groups. This mechanism is distinct from—but ultimately similar in consequence to—intentional barriers such as ID requirements or registration restrictions. In line with Besley and Persson (2010) and Acemoglu et al. (2020), I view election administration as part of the public service infrastructure through which the state signals competence, inclusion, and respect. When that infrastructure fails, it alters beliefs about procedural fairness and lowers participation well beyond the affected election. Heterogeneity analyses show that declines are steepest in precincts with higher shares of young voters, welfare recipients, and residents with migration backgrounds. These are voters whose civic engagement is already fragile—and who appear most susceptible to scarring when participation costs spike unexpectedly.

To probe mechanisms, I link my administrative panel to survey data from the German Longitudinal Election Study (GLES). Immediately after the 2021 election, survey respondents in Berlin report both lower perceived procedural fairness and a higher likelihood of last-minute abstention relative to respondents elsewhere. These results support a dual-mechanism framework: disruptions weaken civic habits and also trigger informational updating about state competence. Fujiwara et al. (2016), Gerber et al. (2003), and Coppock and Green (2016) model voting as habit-forming behavior, while Meredith et al. (2009) and Shino and Smith (2018) show that early disruptions can inhibit routine formation. Separately, work on belief updating in public service delivery (Alsan and Wanamaker, 2018; Lowes and Montero, 2021; Akhtari et al., 2022) finds that bureaucratic failure can reduce institutional trust. My results echo this pattern in the electoral domain.

This paper contributes to three strands of literature. First, it extends work on the economics of voting by showing that *how* elections are administered—not just whether they are legally fair—affects long-run turnout (Cantoni, 2020; Alipour and Lindlacher, 2025; Hodler et al., 2015; Burden et al., 2014; Gerber et al., 2013). Second, it adds to behavioral political economy by highlighting habit disruption and belief decay as causal pathways linking public service failures to disengagement (Bechtel et al., 2018; Kaplan and Yuan, 2020; Olawole, 2023). Third, it speaks to the political economy of state capacity, showing how breakdowns in electoral logistics resemble failures in other public services such as health, social services, or education (Herron and Smith, 2012; Toral, 2023).

Finally, the results have direct policy implications. The costs of administrative failure are not evenly distributed: turnout losses are largest in precincts with many young voters, welfare recipients, and residents with migration backgrounds, i.e. groups that already have weaker formal representation. This is a distributive failure of the state, not just a logistical inconvenience. It is also avoidable. Standard capacity safeguards—maintaining ballot stock buffers, monitoring queue length in real time, deploying mobile backup staff and materials, and scaling postal and early-voting infrastructure where strain is anticipated—are straightforward to implement.

Ensuring that elections run without breakdowns is therefore not merely a technical objective of electoral management. It is part of guaranteeing equal access to democratic voice.

The remainder of this paper proceeds as follows. Section 2 describes the institutional setting, data sources, and the empirical strategy. Section 3 reports the main results. Section 4 explores political consequences. Section 5 presents the conceptual framework. Section 6 concludes.

2 Institutional Setting, Data, and Empirical Strategy

2.1 Institutional Setting

Modern elections are not only exercises in democratic choice but also complex logistical undertakings. Their effectiveness as public services hinges on competent state capacity. This section summarizes the institutional and empirical context for Berlin's 2021 election failures, the subsequent legal rulings, and the data infrastructure used to study their effects.

Berlin offers a valuable case for studying the consequences of administrative failure in a high-capacity democratic state. Germany is widely regarded as having strong institutions, high procedural trust, and a professional civil service. The electoral system is well-established, with routine federal, state, and district-level contests conducted under a uniform legal framework. That such large-scale electoral failure occurred in this setting underscores the potential fragility of even well-functioning democracies when administrative delivery falters.

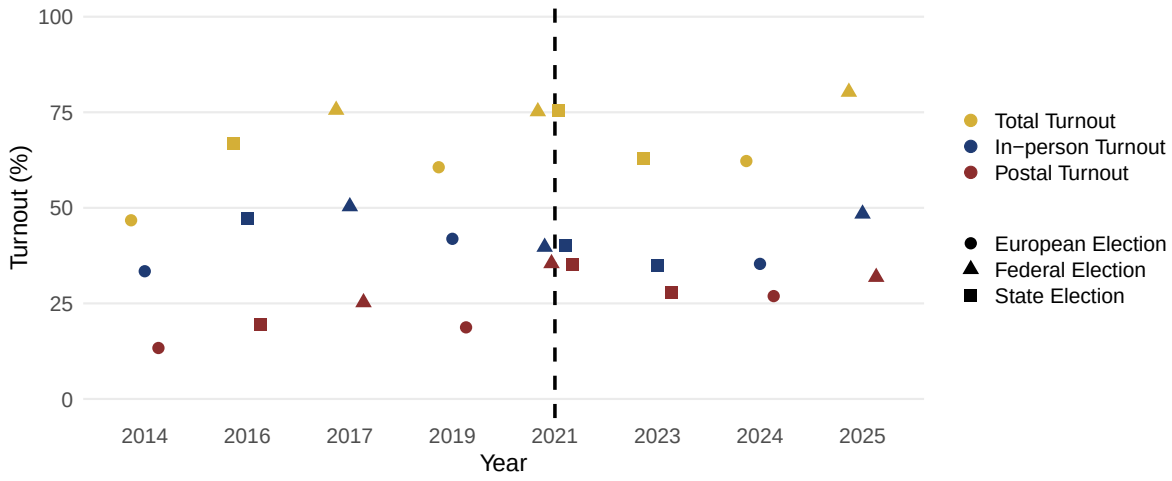
On 26 September 2021, Berlin held four concurrent elections: federal (*Bundestag*), state (*Abgeordnetenhaus*), district councils (*Bezirksverordnetenversammlung*), and a city-wide referendum. These elections took place under COVID-19 precautions and coincided with the Berlin Marathon, which restricted traffic flow and added further stress to the voting process. The result was a cascade of irregularities: ballot shortages, misallocated or missing ballots, polling stations opening late or closing early, and queues exceeding two hours. Media reports and citizen complaints described widespread dysfunction across the city.

These failures triggered extensive legal review. The Berlin Constitutional Court annulled the entire state and district elections, citing "systematic, city-wide electoral disruptions" that had made a proper conduct of the election "impossible in broad parts of the territory" (Verfassungsgerichtshof des Landes Berlin, 2022, VerfGH 154/21). The court documented dozens of violations, including ballots for the wrong district, missing materials, overcrowded polling stations, and unlawful closures. In a separate ruling, the Federal Constitutional Court annulled the federal election in 431 precincts, citing "mandatsrelevante Wahlfehler"—errors severe enough to plausibly affect seat allocation in the Bundestag (Bundesverfassungsgericht, 2023, BVerfG 2 BvC 4/23). These included ballot delivery failures, voting after 6:00 p.m. in violation of federal law, and prolonged queuing that exceeded the court's threshold of tolerable burden for voters.

These rulings offer a rare source of court-validated, precinct-level treatment assignment. Unlike studies relying on media reports, perceptions, or self-reported experience, my setting defines treatment based on formal judicial determinations of procedural failure. This ensures that the classification of affected precincts is exogenous to observed turnout or political behavior.

The irregularities constituted a multidimensional disruption to voting. Ballot unavailability and long queues raised participation costs; procedural confusion and unauthorized polling closures likely reduced trust in the electoral process; and road closures due to the marathon imposed physical access frictions. Prior research shows that even minor logistical barriers—such as relocating a polling place or increasing distance to it—can reduce turnout, particularly among marginal voters (Alipour and Lindlacher, 2025; Cantoni, 2020). The Berlin case, by contrast, involves severe, visible breakdowns in electoral delivery in a setting otherwise known for bureaucratic order, providing a test of behavioral and informational mechanisms under extreme but non-strategic administrative stress.

Figure 1: Turnout by Election Type in Berlin since 2014



Notes: The figure presents total (yellow), in-person (blue), and postal turnout (green) for the European (circle), Federal (triangle), and State (square) elections in the sample. The dotted vertical line marks the 2021 election, which triggered the constitutional review.

2.2 Data

Importantly, Germany's electoral system allows for postal voting without excuse, which has steadily expanded since 2008. Voters may request ballots in advance and return them by post or by deposit. As shown in Figure 1, postal turnout has increased over time and now constitutes a major channel of participation. This allows me to distinguish whether discouraged in-person voters substituted toward mail ballots, or dropped out of the electorate entirely. The former implies a shift in vote mode; the latter suggests behavioral disengagement.

I construct a precinct-level panel of electoral returns across nine State, European, and Federal elections between 2014 and 2025. The unit of observation is the postal precinct (*Briefwahlbezirk*), the smallest level for which mode-specific turnout is reported. To allow consistent comparisons over time, all data are harmonized to the 2021 precinct geography using population-weighted

overlays based on 100-meter census grid data (see Appendix B for details on data sources, harmonization, and treatment coding).

My analysis focuses on the 431 precincts invalidated by the Federal Constitutional Court, as these featured high-salience, court-verified disruptions and were subject to a court-ordered rerun of the federal vote in 2024.[1] Precincts affected only by the state court ruling are excluded from the main analysis, as their disruptions (e.g., ballot misallocation without delays) were less visible, and the 2023 state rerun was widely perceived as low-stakes. Including both types of precincts yields qualitatively similar but attenuated estimates (see Appendix A).

Figure 2 shows the geographic distribution of affected postal precincts. While disruptions cluster in inner-city districts, affected postal precincts are present across most boroughs, supporting comparisons with a broad control group.

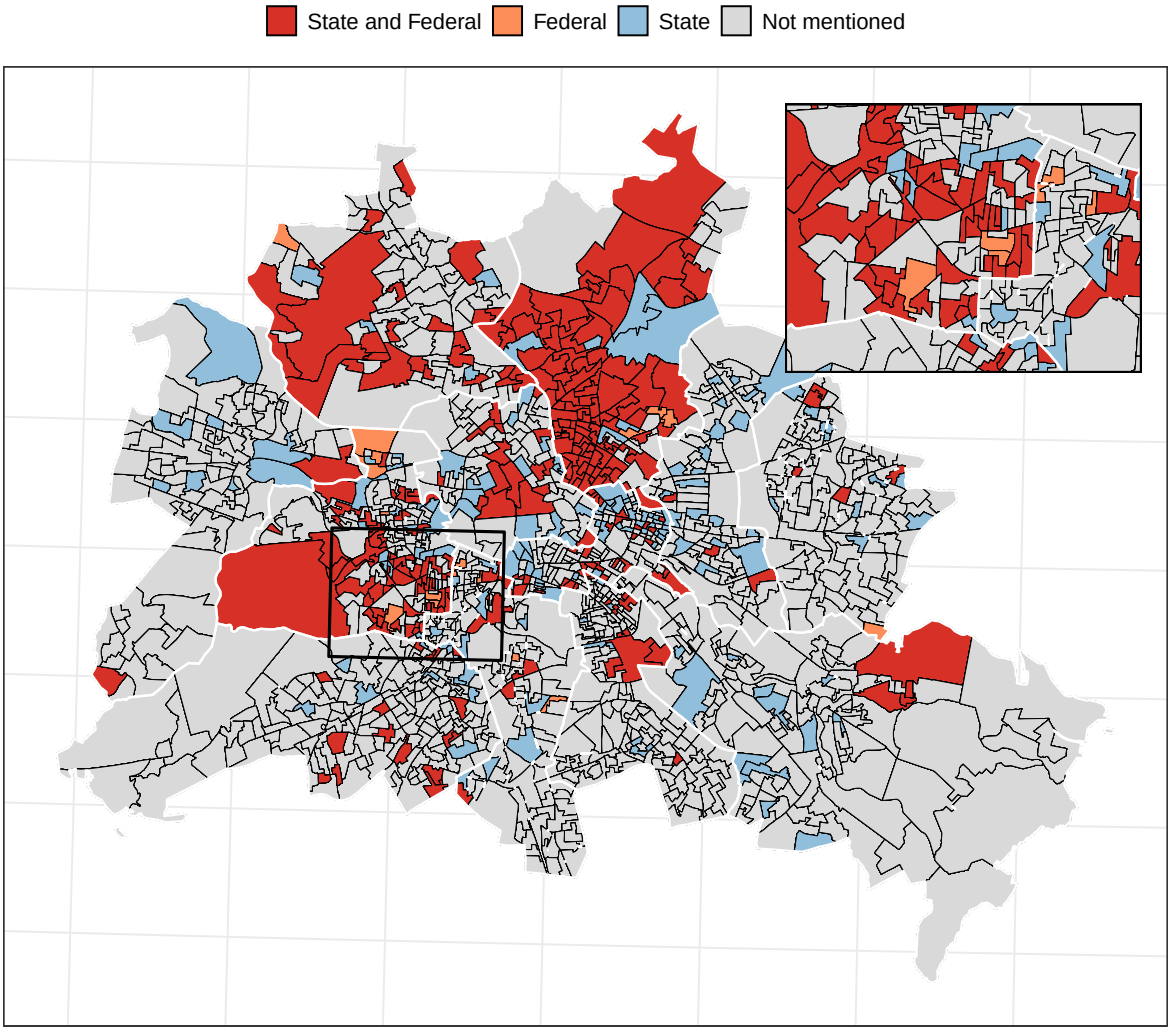
2.3 Identification

I define $Treatment_i = 1$ for any postal precinct i that appears on the Court's annulment list. I define control postal precincts within Berlin as those that were administered under the same election, on the same day, by the same city, but were not annulled. Treated precincts are those for which the Federal Constitutional Court annulled the federal election result from September 26, 2021. The Court annuls an election result only if it documents concrete violations of electoral procedure that are plausibly outcome-relevant ("*mandatsrelevante Wahlfehler*"), e.g. (i) voters receiving the wrong ballot papers, (ii) ballot shortages that forced polls to close temporarily or turn voters away, (iii) queue lengths that were disproportionate or extended well beyond the legal closing time, and (iv) in some cases, voting after 18:00 in violation of federal election law. The Court's determination is procedural, not political: a precinct is annulled because the administration failed to provide a lawful opportunity to vote, not because turnout was low or because any party contested the local result.

Before the 2021 breakdown, treated precincts do not look like "weak" or disengaged areas. In the last pre-breakdown federal election (2017), turnout in treated precincts averaged 78.5%, compared to 76.0% in control precincts; I see similarly higher baseline turnout in the last pre-breakdown state (2016) and European (2019) elections. Treated and control precincts are also very similar in age structure and migration background, and treated precincts actually have lower welfare receipt (11.4% vs. 13.7%). See Table A7. Treatment status is assigned ex post by the Federal Constitutional Court based on documented administrative failures (ballot shortages, unlawful closures, multi-hour queues), not on turnout or partisan complaints. My identification then comes from within-precinct changes after 2021, with precinct fixed effects and district-by-election fixed effects. In Section X I show that, within each election family, treated and control precincts follow parallel turnout dynamics before 2021, and only diverge afterwards.

[1] While these precincts were subject to a court-ordered rerun of the federal election in 2024, I do not interpret the rerun as a separate treatment. The rerun was held more than two years after the original disruption, with minimal campaign activity and very low turnout. It is unlikely to have restored voting habits or trust; see Appendix Appendix B for discussion.

Figure 2: Berlin Precincts by Court Mention



Notes: The figure displays Berlin's postal precincts (*Briefwahlbezirke*) classified by whether they were formally designated as affected by administrative irregularities during the 2021 elections. Status is based on legal rulings by the *Berlin Constitutional Court* (state election), the *Federal Constitutional Court* (federal election), or both. These designations reflect judicial findings of procedural violations and do not distinguish among types of disruption (e.g., ballot errors, long queues). District boundaries are shown in white. The inset in the top right zooms into a representative high-density area to illustrate the fine-grained spatial distribution of statuses. Its location is exemplary and not of substantive analytical relevance.

A potential concern is that the control precincts are not a "pure" counterfactual, because the 2021 breakdown in Berlin's election administration was widely publicized citywide while polls were still open and in the days that followed. Queues lasting hours, ballot shortages, unlawful polling station closures, and voters being told to return later were not experienced in isolation by a handful of precincts; they became a scandal for the entire city and ultimately triggered a constitutional review and partial annulment of results. This creates the possibility of spillovers: residents of precincts that were not formally annulled were nonetheless exposed to the same information shock about the quality and fairness of the election and could have updated their expectations about the cost of voting in future contests. Such spillovers would tend to depress

turnout in both treated and control precincts after 2021, biasing the difference-in-differences estimates toward zero. In other words, any remaining turnout gap between annulled and non-annulled precincts should be interpreted as a lower bound on the causal effect of direct exposure to severe administrative failure.

2.4 Estimation

I estimate a dynamic difference-in-differences model using a stacked event study design. This approach leverages quasi-experimental variation in exposure to administrative disruptions during the 2021 federal election. All treated precincts were affected simultaneously, and voting outcomes are observed across multiple elections before and after the disruption for the same spatial units.

The estimating equation is:

$$Y_{idt} = \sum_{\tau \neq -1} \beta^{\tau} (\mathbb{1}_{\tau=t} \times \text{Treatment}_i) + \alpha_i + \alpha_{dt} + \varepsilon_{idt}, \tag{1}$$

where Y_{idt} denotes the natural log of turnout (total, in-person, or postal) in postal precinct i, district d, and election year t. The variable Treatment_i is an indicator for whether precinct i was identified as affected by severe electoral disruptions in the Federal Constitutional Court ruling. The indicators $\mathbb{1}_{\tau=t}$ denote relative election year, where $\tau - 1$ corresponds to the 2019 European Parliament election (the pre-treatment baseline), and $\tau \in \{-4, -3, -2, 0/F, 0/S, 1, 2, 3\}$ indexes elections from 2014 to 2025.[2]

The model includes postal precinct fixed effects α_i to absorb time-invariant spatial heterogeneity and district-by-election fixed effects α_{dt} to flexibly capture time shocks varying across districts. Standard errors are clustered at the precinct level to account for serial correlation. The coefficients β^{τ} capture the dynamic treatment effects of exposure to administrative failure. Estimates in $\tau + 0$ reflect the immediate effect on turnout in 2021, while post-treatment coefficients ($\tau > 0$) trace the persistence or attenuation of effects over time. This framework allows me to assess whether administrative failures affect not only whether voters turn out, but also how they vote (by mode), and whether these effects are transitory or sustained. The next section presents the results of this analysis and examines heterogeneity across voting channels and voter subgroups.

3 Results

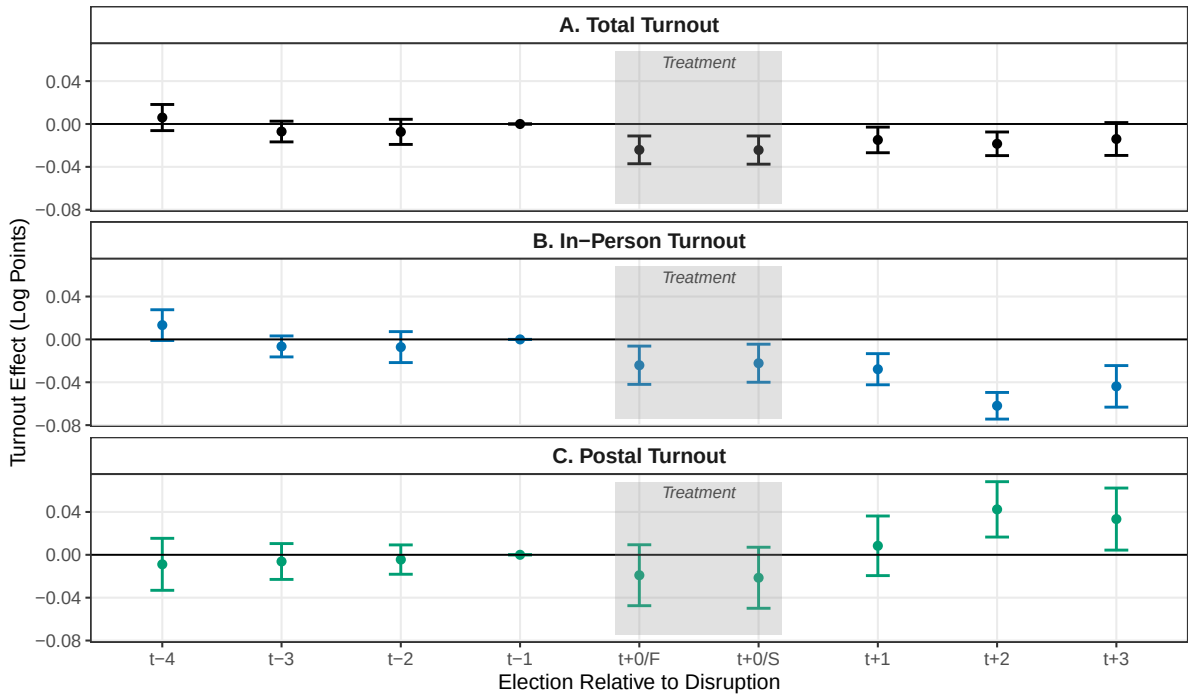
3.1 Effects on Voter Turnout

This section presents the main results from the event study design, which estimates the causal effect of administrative irregularities during the 2021 Berlin State and Federal Elections on voter turnout. Using Equation 1, I track changes in total turnout, in-person voting, and postal voting across treated and untreated precincts over multiple election cycles. Each election is interpreted as a relative period $\tau \in \{-4, -3, -2, -1, 0, +1, +2, +3\}$ with respect to the 2021 disruption.

[2] $0/F$ and $0/S$ refer to the 2021 Federal and State elections, which were held concurrently.

Figure 3 visualizes the estimated coefficients $\hat{\beta}^{\tau}$ by voting mode, and the underlying estimates are reported in Table 1. For interpretability, I convert all log-point estimates into percentage point effects by scaling them with the mean turnout in control precincts in the corresponding election year.

Figure 3: Main Specification



Notes: The figure presents event study results based on Equation 1 for the natural log of postal, in-person, and total turnout. The reference election $(t-1)$ is the 2019 European election. Confidence intervals are drawn at the 95 percent level using standard errors clustered at the postal precinct level. The point estimates and standard errors underlying the results appear in Table 1.

The pre-treatment period serves to assess the plausibility of the parallel trends assumption. Estimated effects in $\tau - 2$ and $\tau - 3$ are close to zero and statistically insignificant across all turnout modes. For instance, total turnout in 2017 differs by just -0.56 percentage points in treated versus control precincts, and postal turnout by -0.12 points. A marginally positive estimate for in-person turnout in 2014 (+0.46 pp, $p = 0.07$) does not persist in later pre-treatment years and is consistent with random variation.

The most substantial effects emerge in the disruption year 2021 ($\tau + 0$). In the Federal election, total turnout in treated precincts declined by 1.82 percentage points relative to controls, from a baseline of 76.4%. This drop is driven entirely by in-person turnout, which fell by 0.95 pp from a baseline of 39.7%. Postal turnout declined only modestly (-0.69 pp) and the estimate is not statistically significant. A similar pattern holds for the State election held on the same day, with total turnout declining by 1.84 pp and in-person turnout by 0.88 pp. Postal turnout again shows no meaningful change. These immediate effects are consistent with the hypothesized cost shock: ballot shortages, long queues, and administrative confusion increased the effort required to vote, particularly for in-person participants.

Two years later, in the court-mandated rerun of the Berlin State election in 2023 ($\tau + 1$), total turnout remains 0.94 percentage points lower in treated precincts, relative to a control group average of 63.8%. This continued suppression is again concentrated in in-person voting (-0.97 pp), while postal turnout rises slightly (+0.24 pp) but remains statistically insignificant. The absence of any compensatory spike suggests a persistent behavioral scarring effect.

By the time of the 2024 European Parliament election ($\tau + 2$), the pattern becomes more pronounced. In-person turnout in treated areas is 2.15 percentage points lower than in controls (baseline: 35.8%), the largest effect observed in any post-treatment period. This decline is partially offset by a significant increase in postal turnout (+1.16 pp), resulting in a net reduction in total turnout of 1.15 pp. These findings suggest gradual behavioral adaptation: some voters exposed to administrative failure appear to shift toward mail voting, but not enough to fully close the turnout gap.

In the 2025 Federal election ($\tau + 3$), in-person turnout remains 2.09 percentage points lower in treated precincts, and total turnout is still depressed by 1.12 points. Postal turnout increases by 1.08 pp, indicating some continued substitution, but the overall participation rate remains below baseline.

Taken together, the estimates show that administrative irregularities exert both immediate and durable effects on electoral participation. The sharp initial drop in in-person turnout aligns with models in which increased participation costs suppress turnout (e.g., Downs, 1957; Pettigrew, 2021). The persistence of these effects—despite the resolution of logistical failures—points to longer-run behavioral scarring and possible belief updating about the electoral process. That the turnout gap remains visible three years later, and across unrelated elections, underscores the potential for even one-time administrative failures to undermine civic engagement.

Finally, I conduct several robustness checks, including alternative fixed effect specifications, placebo treatments in pre-treatment years, and subsample restrictions. These are detailed in Appendix A. The magnitude and statistical significance of treatment effects on in-person turnout are highly stable across specifications, reinforcing the conclusion that the observed effects are not artifacts of model choice or sampling variation.

3.2 Disruption Intensity: Waiting Times

While the main analysis uses a binary treatment indicator based on legal rulings, not all affected precincts experienced irregularities of equal severity. This subsection examines whether the intensity of disruption—measured as maximum waiting time in hours—is associated with differential effects on turnout. I thus shift from a dichotomous to a continuous treatment definition, testing for dose-response effects of election-day failure.

Disruption intensity is operationalized using precinct-level estimates of queue length during peak voting hours, as recorded in the official ruling of the Berlin Constitutional Court. These data, reported in hours, capture the severity of procedural breakdowns at the polling station level and are aggregated to the postal precinct (*Briefwahlbezirk*).[3]

[3] For more detail, see Appendix B.

Table 1: Main Specification

	Total Turnout (1)	In-person Turnout (2)	Postal Turnout (3)
Treatment (t-4)	0.0060	0.0133*	-0.0089
	(0.0062)	(0.0073)	(0.0124)
Treatment (t-3)	-0.0071	-0.0066	-0.0062
	(0.0049)	(0.0050)	(0.0085)
Treatment (t-2)	-0.0073	-0.0072	-0.0044
	(0.0060)	(0.0074)	(0.0070)
Treatment (t+0/Federal)	-0.0241***	-0.0241***	-0.0191
	(0.0066)	(0.0091)	(0.0145)
Treatment (t+0/State)	-0.0243***	-0.0222**	-0.0214
	(0.0067)	(0.0090)	(0.0145)
Treatment (t+1)	-0.0149**	-0.0278***	0.0084
	(0.0061)	(0.0074)	(0.0142)
Treatment (t+2)	-0.0185***	-0.0619***	0.0424***
	(0.0056)	(0.0063)	(0.0132)
Treatment (t+3)	-0.0140*	-0.0438***	0.0333**
	(0.0078)	(0.0099)	(0.0148)
R^2	0.92949	0.86623	0.93030
Observations	12,150	12,150	12,150
Precinct FE	✓	✓	✓
Election-District FE	✓	✓	✓

Notes: The table presents event study results based on Equation 1 for the natural log of postal turnout, in-person turnout, and total turnout. The reference election $(t-1)$ is the 2019 European Election. Standard errors are clustered at the postal precinct level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

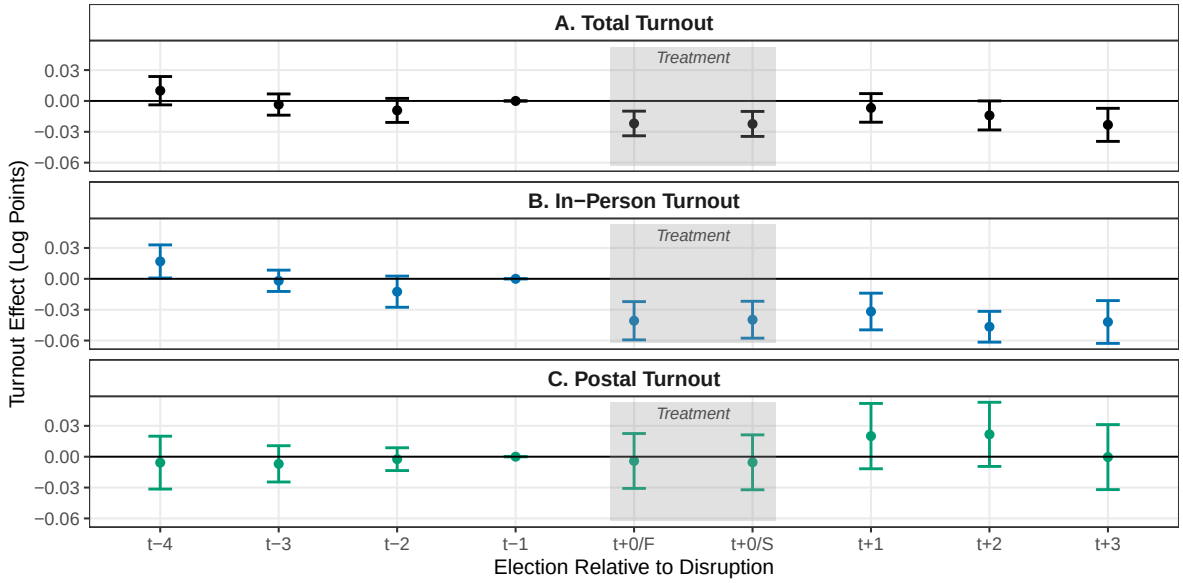
Figure 4 plots the estimated effects using an interaction between waiting time and relative election year, following the stacked event-study framework. The results show a clear and persistent relationship between disruption intensity and turnout suppression.

In the immediate aftermath $(t+0)$, each additional hour of waiting is associated with a 1.36 percentage point decline in total turnout, mirroring the average treatment effect from the binary specification. The effect on in-person turnout is notably stronger: an extra hour of queuing reduces in-person participation by 1.6 to 1.8 percentage points in 2021. These effects remain statistically significant in subsequent cycles. By the 2024 European election $(t+2)$, an additional hour of waiting corresponds to a 4.7 percentage point drop in in-person turnout and a 2.3 point decline in total turnout—exceeding the initial effect at $t+0$.

Postal turnout exhibits the opposite trend: longer wait times in 2021 are associated with small but positive increases in mail voting in $t+2$ and $t+3$, suggesting behavioral adaptation. For example, in precincts that experienced multi-hour queues, postal participation in 2024 rises by up to 2.2 percentage points per hour of prior waiting, partially offsetting in-person losses.

Taken together, these findings reinforce the interpretation that administrative failures depress turnout in a dose-dependent manner. Voters appear to react not only to the presence of irregularities but to their severity. The results also support the substitution narrative: while

Figure 4: Treatment Intensity



Notes: The figure presents triple-interaction estimates based on Equation 1 for the natural log of total turnout, in-person turnout, and postal turnout. The treatment variable is the maximum waiting time (in hours) per postal precinct during the 2021 elections, as recorded by the Berlin Constitutional Court. Each point shows the effect of one additional hour of waiting on turnout in a given election. The sample is restricted to postal precincts that are affected by longer waiting times, disruptions, or longer opening hours, and the control group. The reference election $(t-1)$ is the 2019 European Election. Confidence intervals are drawn at the 95 percent level using standard errors clustered at the postal precinct level. The point estimates and standard errors underlying the results appear in Appendix Table A8. All first and second-order interaction terms required for the identification of the triple-difference estimator are included in the specification or absorbed by the fixed effects.

some affected voters eventually shift to mail voting, this behavioral compensation is incomplete and delayed.

These patterns have both theoretical and policy implications. From a theoretical perspective, the dose-response shape is consistent with models of civic habit scarring and belief updating: more severe disruptions should lead to stronger belief shocks and greater habit discontinuities. From a policy standpoint, the results highlight queue duration as a critical administrative metric. Long lines are not merely logistical frictions; they are visible signals of state dysfunction that can undermine democratic participation in the long run. Monitoring and mitigating wait times—via staffing buffers, queue management systems, or expanded early/postal voting—should be viewed as core components of electoral resilience.

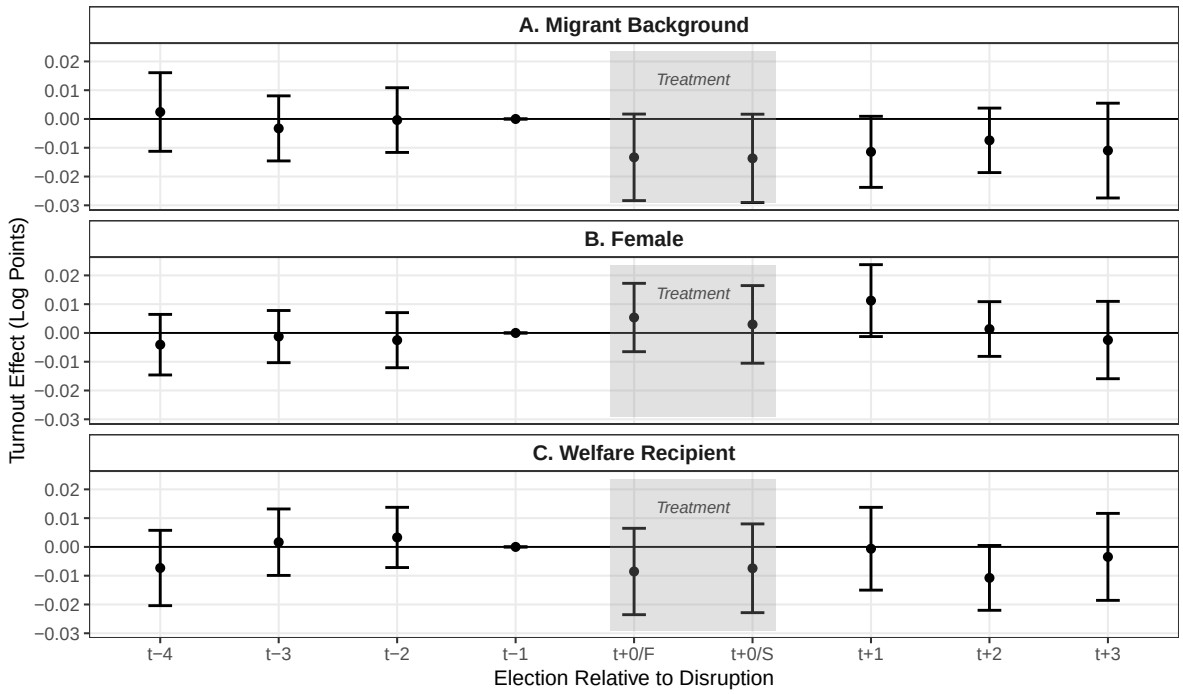
3.3 Heterogeneous Effects by Demographics and Prior Behavior

While the average treatment effects show persistent declines in turnout following administrative failure, existing theory and empirical research suggest that such effects may vary across voter groups. Participation costs and voting habits are not evenly distributed across the electorate, and expectations about the integrity of state services may differ systematically by demographic background. This section investigates these differences by interacting treatment with standardized socio-demographic characteristics in a triple-difference event-study specification:

$$Y_{idfst} = \sum_{\tau \neq -1} \gamma^{\tau}(\mathbb{1}_{\tau=t} \times \text{treatment}_i) + \sum_{\tau \neq -1} \mu^{\tau}(\mathbb{1}_{\tau=t} \times \text{treatment}_i \times Z_i) + \sum_{\tau \neq -1} \rho^{\tau}(\mathbb{1}_{\tau=t} \times Z_i) + \eta_i + \eta_{dt} + \epsilon_{idfst}, \tag{2}$$

where Z_i is a standardized (mean-zero, unit-variance) measure of the demographic characteristic of interest. The coefficients μ^{τ} identify whether the treatment effect varies with that characteristic. Results are visualized in Figure 5 and Figure 6.

Figure 5: Heterogeneity by Socio-Demographics



Notes: The figure presents triple-interaction estimates based on Equation 2 the natural log of total turnout. The reference election $(t-1)$ is the 2019 European Election. Confidence intervals are drawn at the 95 percent level using standard errors clustered at the postal precinct level. The point estimates and standard errors underlying the results appear in Appendix Table A9. All first and second-order interaction terms required for the identification of the triple-difference estimator are included in the specification or absorbed by the fixed effects.

Migrant Background. Precincts with higher shares of adult citizens with a migration background experience systematically larger turnout declines after the 2021 disruption. A one standard deviation increase in migrant share is associated with an additional drop of around 0.5 percentage points at $t+0$, and this differential persists through later elections. This is consistent with prior evidence that political behavior among migrant-background voters is more sensitive to signals of institutional competence and inclusion (Olawole, 2023). If civic habits are less entrenched and trust in electoral administration more fragile in these populations, then visible administrative breakdowns may be particularly demobilizing. These findings also resonate

with broader research on state capacity and participatory inequality (Alsan and Wanamaker, 2018; Akhtari et al., 2022).

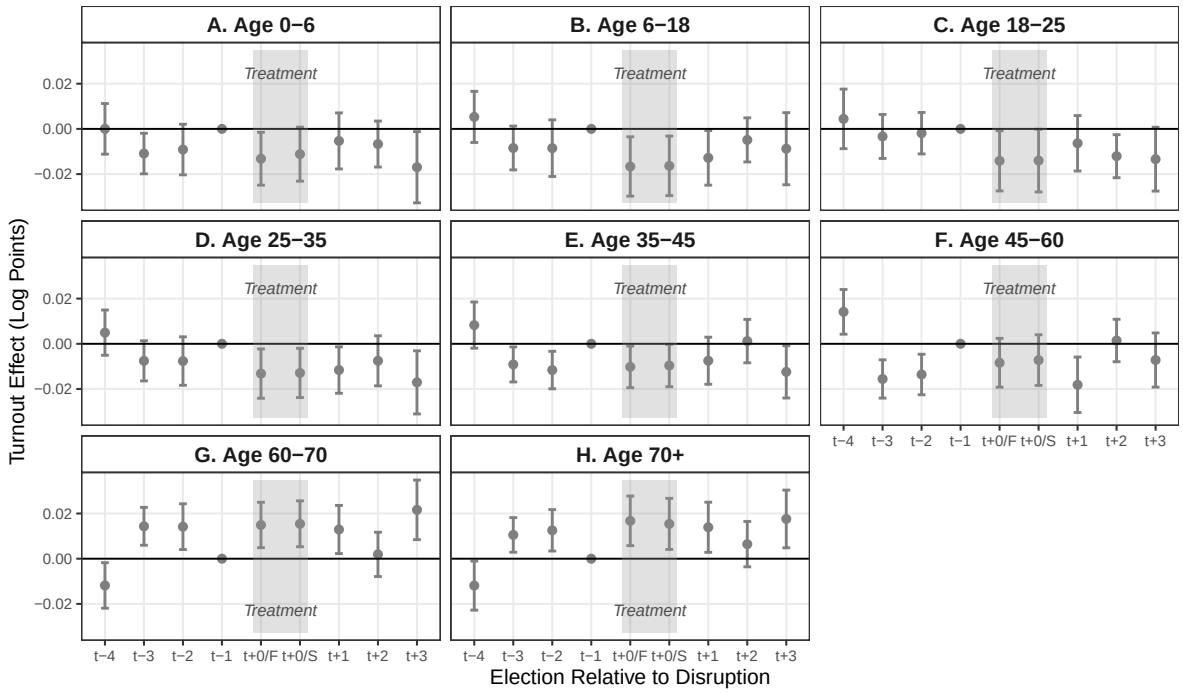
Welfare Recipients. I find similarly persistent effects in precincts with higher shares of welfare recipients. Here, however, the treatment interaction grows in magnitude over time: the effect is small at $t + 0$ but becomes statistically significant by $t + 1$ and peaks in $t + 2$. This delayed dynamic suggests a compounding disengagement process. Voters with lower economic security may initially retain their voting behavior, but interpret unresolved administrative failure as a broader signal of exclusion or neglect. These results are in line with the idea that bureaucratic quality shapes political participation not only through direct experience but also through symbolic representation (Lowes and Montero, 2021; Herron and Smith, 2012).

Gender. I find no robust evidence of heterogeneous effects by gender composition. Across all post-treatment elections, interaction coefficients for the share of female voters are small and statistically indistinguishable from zero. This suggests that, in this context, gender does not systematically condition the behavioral response to electoral disruptions. The result is consistent with prior studies showing that gender gaps in turnout are often small or absent in institutional settings with universal enfranchisement and non-strategic barriers to participation.

Age. Age exhibits a strong and intuitive gradient. Young voters aged 18–25 and 25–35 show the largest treatment effects in the immediate and medium-term aftermath of the disruption, with declines exceeding 1.5 percentage points in some cycles. By contrast, older voters (60+) exhibit no turnout loss, and in later years even show slightly positive deviations. These patterns align with models of civic habit formation: older voters likely possess more stable voting routines, while younger cohorts—many of them new or recent entrants to the electorate—are more vulnerable to disruptions that prevent habit consolidation (Fujiwara et al., 2016; Shino and Smith, 2018; Coppock and Green, 2016). Moreover, younger voters may be more responsive to salient signals about institutional quality, especially in their early voting experiences.

Taken together, these patterns reinforce the interpretation that the long-run effects of administrative failure are not uniform, but concentrated among politically vulnerable or institutionally peripheral populations. The magnitude and timing of turnout decline across subgroups is consistent with mechanisms of disrupted habit formation and revised expectations about state reliability. While I cannot fully disentangle these channels in the data, the heterogeneity results support a behavioral interpretation of electoral disengagement, whereby early disruptions shape subsequent civic trajectories.

Figure 6: Heterogeneity by Age



Notes: The figure presents triple-interaction estimates based on Equation 2 the natural log of total turnout. The reference election $(t-1)$ is the 2019 European Election. Confidence intervals are drawn at the 95 percent level using standard errors clustered at the postal precinct level. The point estimates and standard errors underlying the results appear in Appendix Table A10. All first and second-order interaction terms required for the identification of the triple-difference estimator are included in the specification or absorbed by the fixed effects.

4 Political Consequences of Electoral Disruption

4.1 Behavioral and Attitudinal Effects

In this section, I provide survey-based evidence on the channels that link administrative breakdown in 2021 to persistent declines in electoral participation. My turnout results show that directly affected precincts experience a large and long-lasting reduction in in-person voting, with only gradual substitution toward postal ballots, and little sign of full recovery several years later. These reduced-form patterns are consistent with two mechanisms: (i) a disruption of voting habits, in which being blocked or discouraged from voting in 2021 weakens future turnout propensity, and (ii) an update in beliefs about the cost, fairness, and competence of election administration. I cannot separately identify these channels using precinct-level turnout alone. I therefore turn to individual-level data that speak to both mechanisms.

I use pooled data from the German Longitudinal Election Study (GLES) Rolling Cross-Section surveys for the 2017, 2021, and 2025 federal elections. I estimate a difference-in-differences specification that compares respondents in Berlin (the treated city) to respondents in the rest of Germany, using 2017 as the pre-treatment baseline:

$$Y_{ist} = \sum_{\kappa \neq 2017} \beta^{\kappa} \left(\mathbb{1}_{\kappa=t} \times \text{Berlin}_i \right) + \gamma_s + \lambda_t + u_{ist}, \tag{3}$$

where Y_{ist} is the outcome of interest for individual i in state s and election year t, γ_s are state fixed effects, and λ_t are year fixed effects. Standard errors are clustered at the state level. I interpret the coefficients β^{κ} as the differential change in Berlin, relative to the rest of Germany, in election-year κ compared to 2017.

Panel A of Figure 7 reports effects on beliefs about procedural integrity. I measure whether respondents agree that "the federal election was conducted correctly and fairly." Immediately after the 2021 election, Berlin respondents are 9 to 13 percentage points less likely to affirm that the election was conducted properly, relative to respondents elsewhere in Germany. This sharp drop is consistent with an informational shock: visible ballot shortages, multi-hour queues, temporary polling station closures, and instances of voting after the legal closing time plausibly sent a strong negative signal about the state's administrative capacity. By 2025, the Berlin–rest-of-Germany gap in perceived electoral fairness narrows substantially, which suggests that stated trust in procedural integrity partially recovers over time. However, Berlin respondents continue to express somewhat lower confidence than respondents in other states, indicating that the perception of administrative quality does not fully revert.

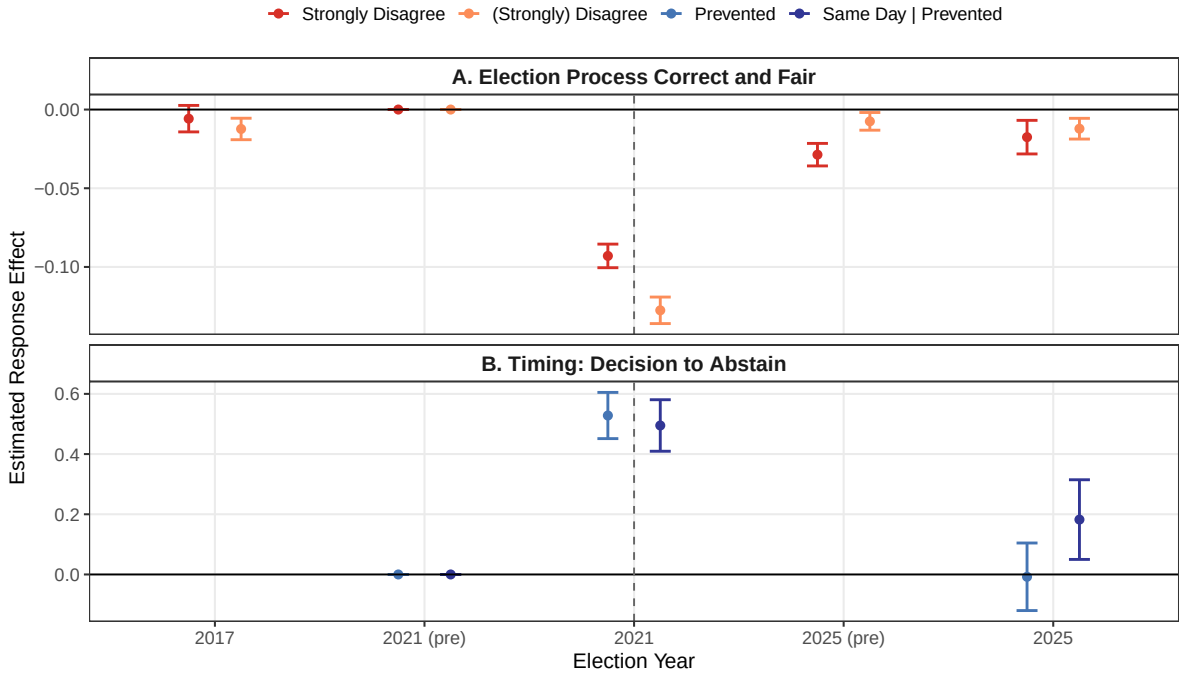
Panel B focuses on abstention timing, using a subsample of roughly 510 self-identified nonvoters. Here I ask *when* abstention was decided and *why* respondents did not cast a ballot. In 2021, nonvoters in Berlin are dramatically more likely than nonvoters elsewhere to report that they decided not to vote on election day itself or that they were effectively prevented from voting. The point estimates for these differences are large — on the order of 49 to 53 percentage points — and point to acute, same-day frictions at the polling place rather than long-run disengagement or lack of political interest. In other words, Berlin abstainers disproportionately describe last-minute breakdowns in access, not gradual loss of motivation.

I interpret these two sets of results as jointly informative about the behavioral channels behind the persistent turnout effects in treated precincts. First, the spike in reported day-of or forced non-participation in Berlin in 2021 is exactly what I would expect if administrative failure interrupted the act of voting for marginal participants. When a voter is willing to vote but is delayed for hours, told to return later, or encounters a polling station that has run out of the correct ballots, that voter is at high risk of dropping out *today*. Because turnout is habit-forming, that missed act of voting can depress participation in subsequent elections, even after the immediate disruption has passed. This mechanism directly links an acute logistical breakdown in 2021 to lower turnout years later.

Second, the sharp decline in perceived electoral fairness in Berlin in 2021 is consistent with an expectations channel: voters update their beliefs about how costly, disorganized, or procedurally unfair future in-person voting will be. This belief channel raises the expected cost of voting in subsequent elections. The fact that these stated perceptions partially rebound by 2025, while the turnout gap in affected precincts does *not* fully close, suggests an important asymmetry: beliefs about fairness can heal faster than voting behavior does. That is, even after I observe recovery in stated trust, I continue to observe depressed in-person turnout in previously disrupted precincts. This pattern is consistent with habit disruption having a durable behavioral footprint.

Figure 7 summarizes these results.

Figure 7: Behavioral and Attitudinal Responses in Berlin vs. the Rest of Germany



Notes: I plot coefficients from Equation 3 using GLES Rolling Cross-Section data from the 2017, 2021, and 2025 federal elections. Panel A shows agreement with the statement that the election was conducted correctly and fairly. Panel B is restricted to self-reported nonvoters (approximately 510 individuals across years) and reports the probability that respondents either (i) decided not to vote on election day or (ii) state that they were prevented from voting. For interpretability, the y-axis in each panel can be read as the difference (in percentage points) between Berlin and the rest of Germany in a given election year, relative to 2017. Standard errors are clustered at the state level. The reference category is 2017. The point estimates and standard errors underlying the results appear in Appendix Table A11.

I emphasize two caveats. First, Berlin sample sizes in the GLES are modest, especially once I condition on nonvoting. The large point estimates in Panel B therefore come with non-trivial uncertainty. Second, the difference-in-differences design in Equation 3 leverages pre/post comparisons and cross-state contrasts, but it does not provide experimental variation. I therefore treat these results as descriptive evidence consistent with the mechanisms in my conceptual framework, rather than as an independent causal research design.

With those caveats in mind, the survey patterns are consistent with the turnout evidence. Immediately after the breakdown, Berliners both report sharply lower confidence in electoral administration and disproportionately describe last-minute or forced abstention. Years later, stated trust in electoral conduct partially normalizes, but turnout in affected precincts remains lower, particularly for in-person voting. This time profile supports a two-step story: an acute administrative shock in 2021 both blocks some voters from participating that day and teaches them that voting is costly, and the resulting disruption to their voting habit persists even after beliefs about procedural fairness begin to recover.

4.2 Administrative Response and Institutional Learning

The observed disruptions in 2021 raise the question of whether electoral administrators recognized the failures and adjusted the electoral infrastructure in response. If long queues and procedural breakdowns were interpreted as a failure of local election delivery, I might expect compensatory efforts in subsequent cycles—such as reallocating or expanding polling station capacity. This subsection examines whether precincts that experienced more severe disruption were more likely to receive additional polling places in later elections.

I estimate difference-in-differences regressions in which the outcome is the number of polling stations in a given precinct and election. The key independent variable is the intensity of disruption in 2021, measured as the total reported queueing time (in hours) across all in-person polling stations within a postal precinct. This variable is interacted with election year indicators for 2023, 2024, and 2025. The specification includes precinct and election fixed effects. Results are shown in Table 2.[4]

I find evidence of a delayed institutional response. In 2024, each additional hour of wait time in 2021 is associated with an increase of approximately 0.480 polling stations per precinct, significant at the 5% level. A positive but smaller relationship appears in 2025 (0.093 polling stations per additional hour of waiting time). No statistically significant change is detected in 2023, when the Berlin state rerun occurred. This timing pattern suggests that administrative adjustments were not immediate but instead reflect a lagged response, consistent with internal review cycles, audit-triggered corrections, or reputational incentives to prevent recurrence in higher-salience contests.

Although I cannot directly assess whether the observed capacity expansion mitigated turnout decline, these findings offer suggestive evidence of bureaucratic learning. The allocation of additional polling infrastructure appears targeted to previously disrupted areas, indicating that electoral administrators responded to performance failures by adjusting resource deployment.

From a policy perspective, these results underscore that electoral institutions are not static. While often modeled as exogenous rule-makers or neutral implementers, electoral bureaucracies in high-capacity democracies may exhibit adaptive behavior—especially after highly visible breakdowns. Future work could examine whether these institutional responses closed participation gaps, either by reducing wait times or by restoring confidence among affected voters.

4.3 Downstream Electoral Composition

While the primary effect of administrative failure is reduced participation, disruptions may also alter the partisan composition of the electorate. If certain parties rely more heavily on politically marginal voters—or if specific groups update their partisan preferences in response to institutional breakdowns—then turnout shocks may produce measurable shifts in vote shares. This subsection examines whether treated precincts exhibit systematic changes in party support following the 2021 disruptions.

[4] Administrative data on polling station counts is available from 2021 through 2025.

Table 2: Administrative Response: Change in Polling Station Capacity

	Number of Polling Stations	
	(1)	(2)
Treatment (t+1)	-0.0171	0.0407
	(0.0260)	(0.0321)
Treatment (t+2)	0.8199***	0.4798***
	(0.1918)	(0.1697)
Treatment (t+3)	0.0566	0.0927***
	(0.0351)	(0.0336)
Specification	Binary	Waiting Time in h
R^2	0.89819	0.89723
Observations	5,400	5,400
Precinct FE	✓	✓
Election FE	✓	✓

Notes: The table presents event study results based on a similar regression as Equation 1 for the number of polling stations within a postal precinct. The reference election $(t + 0)$ is the 2021 Federal Election. Standard errors are clustered at the postal precinct level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

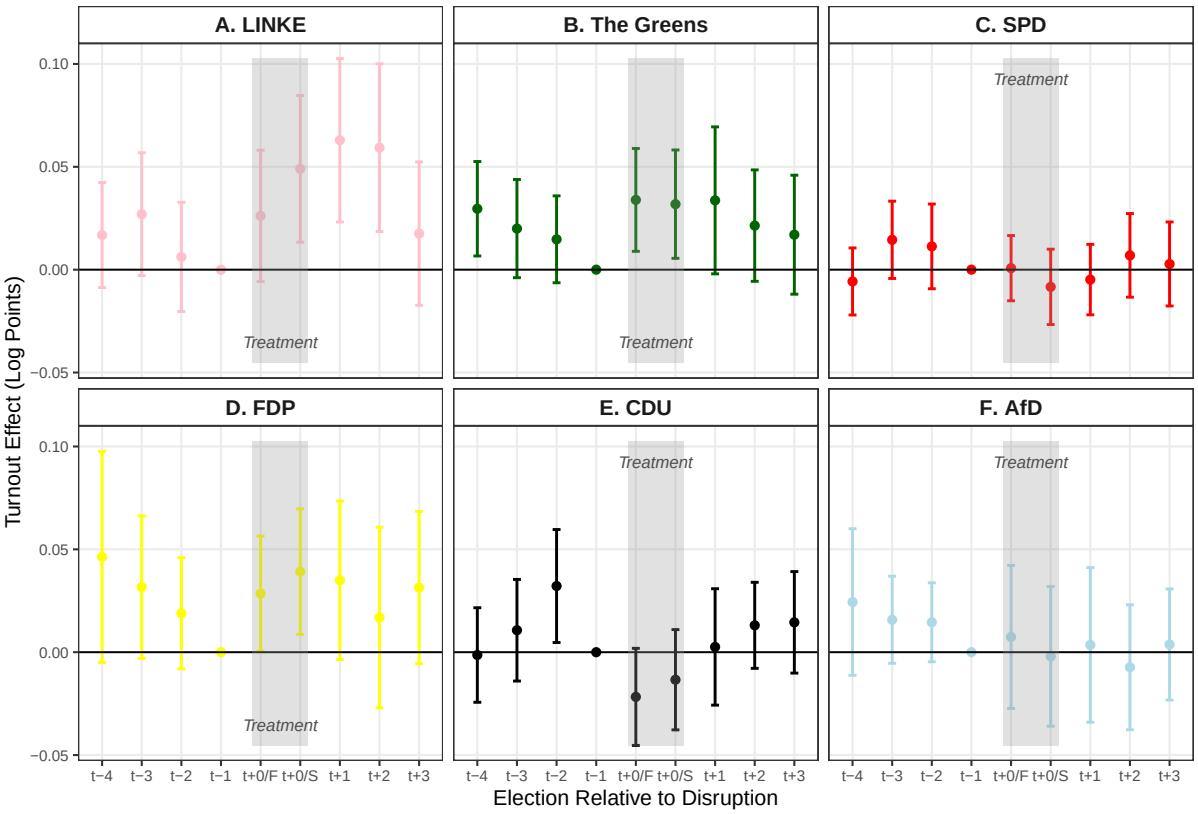
I estimate treatment effects on party vote shares using the same stacked event-study specification as in the main turnout analysis, including postal precinct and election-district fixed effects. Figure 8 summarizes results across four pre-treatment and three post-treatment elections. Most pre-treatment estimates are small and statistically indistinguishable from zero, consistent with the identifying assumption of parallel trends.

In the immediate post-treatment period $(t + 0)$, I observe modest shifts in the distribution of party support. CDU vote shares decline by up to 0.8 percentage points, while the FDP and LINKE gain approximately 0.2–0.4 and 0.6 percentage points, respectively. These trends persist into $t + 1$, where the CDU registers a statistically significant loss of 0.85 points. Vote shares for SPD, Greens, and AfD show no consistent treatment effects across post-treatment elections.

While these partisan shifts are relatively small compared to overall turnout losses, they are directionally consistent with theoretical expectations. To the extent that electoral mismanagement triggers belief updating about state competence, voters with low institutional trust may be more likely to disengage or switch support. CDU losses may reflect reputational costs associated with governing responsibility, particularly if voters perceived the party as accountable for administrative oversight. FDP gains may reflect a substitution effect among more state-skeptical voters. The increase in LINKE vote shares is less easily explained, but may indicate mobilization in response to perceived institutional marginalization or as a protest vote.

These results suggest that administrative failures can affect not only who votes but also the partisan landscape in which votes are cast. While the magnitude of vote-share shifts is modest, the fact that effects persist across multiple elections underscores the potential for even non-strategic failures in service delivery to influence democratic representation.

Figure 8: Election Results (Vote Share)



Notes: The figure presents event study results based on Equation 1 for the vote share of the political parties represented in the federal parliament *LINKE* (Left), *The Greens*, the *SPD* (Social Democrats), the *FDP* (Liberals), the *CDU* (Christian Conservatives), and the *AfD* (right-wing populist). The reference election $(t-1)$ is the 2019 European election. Confidence intervals are drawn at the 95 percent level using standard errors clustered at the postal precinct level. The point estimates and standard errors underlying the results appear in Appendix Table A12.

5 Theoretical Interpretation: Habit Scarring and Belief Decay

Why do one-off administrative failures lead to persistent declines in voter turnout, even after conditions normalize? I present a simple behavioral framework in which disruptions affect future participation through two reinforcing channels: weakened civic habits and revised expectations about the cost of voting. This helps explain why even a transient failure in electoral service delivery can produce long-run changes in turnout, especially among groups with fragile engagement.

A voter's decision to turn out at time t is governed by a utility function:

$$U_t = p_t B + d_t - c_t, \tag{4}$$

where $p_t B$ reflects the perceived benefit of voting (instrumental or expressive), d_t captures internalized civic motivation or habit, and c_t is the anticipated cost of participation. Voting occurs if $U_t > 0$.

Disruptions—such as long queues, confusion, or ballot shortages—raise c_t and can induce abstention even among otherwise willing voters. If voting is habit-forming, as suggested by Fujiwara et al. (2016) and Coppock and Green (2016), this abstention has dynamic consequences. I model habit formation as:

$$d_{t+1} = \bar{d} + \rho v_t, \tag{5}$$

where $v_t \in \{0, 1\}$ is the turnout decision and $\rho > 0$ captures reinforcement. Abstention in t reduces d_{t+1}, making future turnout less likely. Because elections are infrequent, this habit loss may persist across cycles, particularly for new or marginal voters.

Disruptions may also affect expectations. Let $s = \theta + \varepsilon$ denote a noisy signal about the competence of electoral administration, where θ is true service quality. A negative signal—such as visible failure on election day—induces belief updating. I model perceived cost at $t + 1$ as:

$$c_{t+1} = \bar{c} + \phi(s), \quad \phi'(s) < 0, \tag{6}$$

where $\phi(s)$ captures increased perceived costs due to pessimistic expectations. Even if service quality recovers, the memory of prior failure can raise c_{t+1}, reducing turnout. This mechanism aligns with evidence on expectation-driven voting behavior (Kawai et al., 2021) and belief formation from public service encounters (Alsan and Wanamaker, 2018; Lowes and Montero, 2021; Akhtari et al., 2022).

Taken together, this framework implies that turnout at $t + 1$ depends not just on structural factors, but also on prior experience. A single disruption can decrease turnout both by breaking behavioral reinforcement and by inducing belief-based cost inflation. These dynamics are particularly relevant for groups whose voting is less routinized or more sensitive to trust shocks—such as young voters, migrants, or welfare recipients. The model thus offers a plausible explanation for the persistence and concentration of turnout declines observed in the data.

Empirically, I cannot fully disentangle the two mechanisms, but my results are consistent with their joint operation. Heterogeneity analyses show stronger turnout effects among groups likely to have weaker civic routines or higher institutional sensitivity. Moreover, survey data from the GLES suggest that treated voters were more likely to abstain at the last minute and reported lower confidence in electoral fairness—patterns consistent with both habit disruption and belief revision. While not conclusive, this evidence supports the interpretation that administrative failure can disrupt civic behavior through both experiential and informational channels.

My approach extends the classical calculus of voting (Riker and Ordeshook, 1968) by incorporating two dynamic channels: behavioral reinforcement and belief revision. Unlike static-cost models, I treat voting as a learned behavior, shaped by past experience with the state. While prior work has examined legal or institutional barriers to participation, I show that non-strategic failures in electoral administration—when salient—can generate long-lasting behavioral change. This has broader implications for the political economy of public services: even isolated administrative failures may generate negative externalities for democratic inclusion.

6 Conclusion

This paper shows that even non-strategic administrative failures on election day can produce lasting negative effects on political participation. Exploiting quasi-experimental variation from Berlin's 2021 electoral disruptions, I estimate a persistent turnout decline of 1.8 percentage points (2.4%) in affected precincts. These losses are concentrated among young voters, residents with migration backgrounds, and welfare recipients—groups whose civic engagement is already fragile. In-person voting falls sharply, and while some substitution into postal voting occurs over time, it remains incomplete.

Survey evidence suggests that visible failures undermine institutional trust and trigger last-minute abstention, supporting a dual-mechanism framework: disrupted habit formation and belief updating. Heterogeneity analyses further show that voters with lower baseline engagement are most sensitive to such disruptions.

The effects extend beyond participation. Although partisan vote shifts are modest, I document meaningful changes in electoral composition and evidence of administrative learning: polling station capacity expanded in areas previously affected by long queues. These findings underscore that administrative competence is not only a technical issue but a democratic one.

Conceptually, the results bridge behavioral political economy and the literature on state capacity. Elections are recurring encounters with the state, and their quality influences not only what voters do but what they expect. When these encounters fail, however unintentionally, they may disengage the very populations most in need of inclusion.

From a policy perspective, the results are both cautionary and actionable. Administrative failure imposes durable costs on democratic participation, but these costs are avoidable. Low-cost interventions, such as ballot stock buffers, staffing audits, and queue monitoring, could prevent similar breakdowns. As democracies face increasing logistical strain, ensuring that elections run smoothly must be treated as a component of state capacity and a foundation of democratic legitimacy.

Declaration of generative AI and AI-assisted technologies in the manuscript preparation process

During the preparation of this work the author(s) used ChatGPT (OpenAI, San Francisco, CA, USA) in order to improve writing style and clarity and to obtain coding support for data analysis. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article.

References

Acemoglu, D., Cheema, A., Khwaja, A. I., and Robinson, J. A. (2020). Trust in state and nonstate actors: Evidence from dispute resolution in pakistan. *Journal of Political Economy*, 128(8):3090–3147.

Akhtari, M., Moreira, D., and Trucco, L. (2022). Political turnover, bureaucratic turnover, and the quality of public services. *Quarterly Journal of Economics*, 137(4):2385–2432.

Alipour, J.-V. and Lindlacher, V. (2025). No surprises, please: Voting costs and electoral turnout. *Journal of Political Economy Microeconomics*, 3(1):000–000.

Alsan, M. and Wanamaker, M. (2018). Tuskegee and the health of black men. *The quarterly journal of economics*, 133(1):407–455.

Bechtel, M. M., Hangartner, D., and Schmid, L. (2018). Compulsory voting, habit formation, and political participation. *Review of Economics and Statistics*, 100(3):467–476.

Besley, T. and Persson, T. (2010). State capacity, conflict, and development. *Econometrica*, 78(1):1–34.

Bundesverfassungsgericht (2023). Urteil des Zweiten Senats vom 19. Dezember 2023 – 2 BvC 4/23 – Bundestagswahl Berlin – Wahlprüfung. Available at: https://www.bundesverfassungsgericht.de/SharedDocs/Entscheidungen/DE/2023/12/cs20231219_2bvc000423.html.

Burden, B. C., Canon, D. T., Mayer, K. R., and Moynihan, D. P. (2014). Election laws, mobilization, and turnout: The unanticipated consequences of election reform. *American Journal of Political Science*, 58(1):95–109.

Cantoni, E. (2020). A precinct too far: Turnout and voting costs. *American Economic Journal: Applied Economics*, 12(1):61–85.

Coppock, A. and Green, D. P. (2016). Is voting habit forming? new evidence from experiments and regression discontinuities. *American Journal of Political Science*, 60(4):1044–1062.

Downs, A. (1957). An economic theory of democracy. *Harper and Row*, 28.

Fujiwara, T., Meng, K., and Vogl, T. (2016). Habit formation in voting: Evidence from rainy elections. *American Economic Journal: Applied Economics*, 8(4):160–188.

Gerber, A. S., Green, D. P., and Shachar, R. (2003). Voting may be habit-forming: evidence from a randomized field experiment. *American journal of political science*, 47(3):540–550.

Gerber, A. S., Huber, G. A., and Hill, S. J. (2013). Identifying the effect of all-mail elections on turnout: Staggered reform in the evergreen state. *Political Science research and methods*, 1(1):91–116.

Hainmueller, J. (2012). Entropy balancing for causal effects: A multivariate reweighting method to produce balanced samples in observational studies. *Political analysis*, 20(1):25–46.

Herron, M. C. and Smith, D. A. (2012). Souls to the polls: Early voting in florida in the shadow of house bill 1355. *Election Law Journal*, 11(3):331–347.

Heß, S. (2017). Randomization inference with stata: A guide and software. *The Stata Journal*, 17(3):630–651.

Hodler, R., Luechinger, S., and Stutzer, A. (2015). The effects of voting costs on the democratic process and public finances. *American Economic Journal: Economic Policy*, 7(1):141–171.

Kaplan, E. and Yuan, H. (2020). Early voting laws, voter turnout, and partisan vote composition: Evidence from ohio. *American Economic Journal: Applied Economics*, 12(1):32–60.

Kawai, K., Toyama, Y., and Watanabe, Y. (2021). Voter turnout and preference aggregation. *American Economic Journal: Microeconomics*, 13(4):548–586.

Lowes, S. and Montero, E. (2021). The legacy of colonial medicine in central africa. *American Economic Review*, 111(4):1284–1314.

Meredith, M. et al. (2009). Persistence in political participation. *Quarterly Journal of Political Science*, 4(3):187–209.

Olawole, I. M. (2023). The consequences of service delivery for voter turnout in nigeria. *Electoral Studies*, 86:102675.

Pettigrew, S. (2021). The downstream consequences of long waits: How lines at the precinct depress future turnout. *Electoral studies*, 71:102188.

Riker, W. H. and Ordeshook, P. C. (1968). A theory of the calculus of voting. *American political science review*, 62(1):25–42.

Shino, E. and Smith, D. A. (2018). Timing the habit: Voter registration and turnout. *Electoral Studies*, 51:72–82.

Toral, G. (2023). How patronage delivers: Political appointments, bureaucratic accountability, and service delivery in brazil. *American Journal of Political Science*.

Verfassungsgerichtshof des Landes Berlin (2022). Urteil vom 16. November 2022 – VerfGH 154/21 – Wahl zum Abgeordnetenhaus von Berlin 2021. Available at: https://www.verfgh.berlin.de/entscheidungen/verfgh-154-21/.

Online Appendix

Race to the Polls: The Berlin Marathon and its Electoral Consequences
by *Marius Kröper*

A Robustness Checks

A.1 Placebo: Randomization Inference

Table A1: Randomized Inference

	Total Turnout (1)	In-person Turnout (2)	Postal Turnout (3)
Treatment (t-4)	0.0138	0.3645	-0.3507
	(0.2945)	(0.2387)	(0.2275)
	[0.9710]	[0.1021]	[0.1151]
Treatment (t-3)	-0.4513	-0.1428	-0.3085*
	(0.2826)	(0.2038)	(0.1677)
	[0.1502]	[0.5105]	[0.0641]*
Treatment (t-2)	-0.3090	-0.2065	-0.1025
	(0.3131)	(0.3027)	(0.1634)
	[0.3333]	[0.5125]	[0.5726]
Treatment (t+0/Federal)	-1.390***	-1.122***	-0.2684
	(0.3503)	(0.3873)	(0.4113)
	[0.0000]***	[0.0040]***	[0.5205]
Treatment (t+0/State)	-1.437***	-1.047***	-0.3899
	(0.3671)	(0.3861)	(0.4172)
	[0.0000]***	[0.0100]**	[0.3413]
Treatment (t+1)	-0.8260**	-1.162***	0.3361
	(0.3884)	(0.2949)	(0.3725)
	[0.0430]**	[0.0000]***	[0.4154]
Treatment (t+2)	-1.143***	-2.317***	1.173***
	(0.3778)	(0.2601)	(0.3695)
	[0.0040]***	[0.0000]***	[0.0040]***
R^2	0.95213	0.89287	0.92054
Observations	9,808	9,808	9,808
Precinct FE	✓	✓	✓
Election-District FE	✓	✓	✓

Notes: The table presents event study results based on Equation 1 for the natural log of postal turnout, in-person turnout, and total turnout. The p-values calculated based on the randomized inference approach following Heß (2017) are shown in brackets. The reference election ($t-1$) is the 2019 European Election. Standard errors are clustered at the postal precinct level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

I utilize a randomization inference approach following Heß (2017), conducting 999 permutations while maintaining the number of treated postal precincts within each district. This method assesses whether the observed results could arise by random chance. The results, presented in Table A1, show significant treatment effects consistent with the original analysis, thus reinforcing the reliability of I findings. This approach helps to ensure that the results are due to the actual treatment effect of the irregularities during the 2021 Federal elections and not random variation.

A.2 Treatment Definition: Federal vs. State Disruptions

While the core empirical design aggregates disruptions into a single treatment indicator, election-day irregularities varied substantially in nature and timing. In particular, disruptions occurred during either the Federal or the State election held concurrently in 2021, enabling an indirect exploration of heterogeneity by election tier. The most severe disruptions occurred during the Federal election and involved highly visible administrative failures: precincts accepting ballots after the legal closing time, redirecting voters between locations, or remaining open well past 6:00 p.m. By contrast, precincts affected only at the State level were typically flagged for procedural violations such as missing or misallocated ballots—issues that may have been less visible to voters or corrected before deterring participation. To probe whether the salience of disruptions matters for turnout effects, I estimate separate event-study specifications for three mutually exclusive groups: (i) precincts affected during the Federal election (I primary treatment group); (ii) precincts affected only during the State election; and (iii) unaffected precincts, which serve as the counterfactual. Figure A1 presents the resulting trends. Turnout suppression is concentrated in precincts disrupted during the Federal election: the immediate decline at $t+0$ is statistically significant and remains directionally negative in subsequent elections. By contrast, precincts with State-only disruptions exhibit no detectable turnout effects, consistent with the view that these disruptions were less salient or behaviorally consequential. This interpretation is supported by descriptive evidence on the nature of irregularities. Table A2 tabulates the frequency of ballot-related errors, operational breakdowns, and violations of legal closing times. Precincts affected during the Federal election were far more likely to experience high-salience disruptions—such as 212 cases of late closing and 101 classified as general disruptions—compared to those affected only at the State level. Taken together, these findings suggest that persistent disengagement is not driven by treatment status, but emerges when administrative failures are both visible and severe enough to undermine voters' perceptions of procedural legitimacy.

Table A2: Treatment Overview

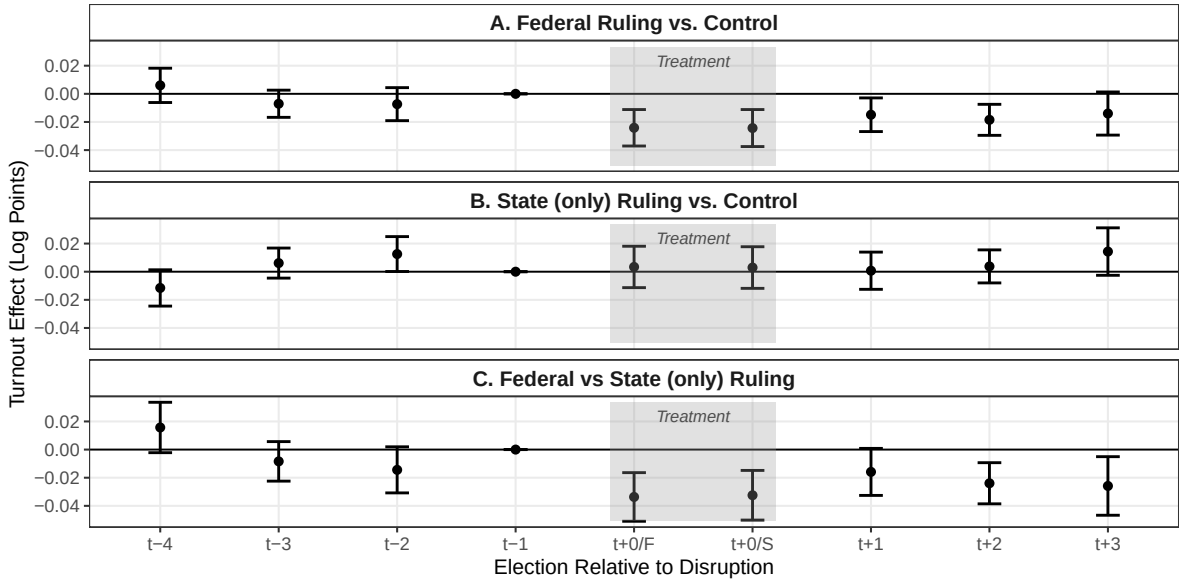
Treatment Federal	Treatment State	Ballot Error	Ballot Missing	Disruption	Unusual Waiting	Open After Official Closing	N
yes	*yes*	32	13	101	90	212	277
no	*yes*	45	61	2	55	1	157
Subsample: Disruption = 0 & Open A.O. Closing = 0							
yes	*yes*	1	2	–	16	–	17
no	*yes*	44	61	–	55	–	154

Notes: The table presents the number of precincts by treatment during the Federal and State elections. The type of irregularity is available only for the State election.

A.3 Controls

In the main specification, I only control for the log number of eligible voters held constant to the 2017 Federal Election, the last federal election before the treatment, and interact them with the

Figure A1: Heterogeneity by Type of Treatment



Notes: The figure presents triple-interaction estimates for the natural log of postal turnout, in-person turnout, and total turnout. The reference election $(t-1)$ is the 2019 European Election. Confidence intervals are drawn at the 95 percent level using standard errors clustered at the postal precinct level.

election indicators. To test the robustness of the main specification, I add further time-invariant controls, holding them constant for 2019. Controls are the share of the following groups: German citizens with migrant backgrounds, foreigners, persons subject to social security, 6-18 year old, 18-25 year old, 65 to 99 year old, EU citizens, persons living in a civil union, Single households; additionally I control for the log number of eligible voters. Results are shown in Table A3

Table A3: Robustness: Controls

	Total Turnout (1)	In-person Turnout (2)	Postal Turnout (3)
Treatment (t-4)	0.0055	0.0129**	-0.0098
	(0.0058)	(0.0065)	(0.0122)
Treatment (t-3)	-0.0049	-0.0043	-0.0040
	(0.0041)	(0.0043)	(0.0080)
Treatment (t-2)	-0.0069	-0.0061	-0.0061
	(0.0043)	(0.0052)	(0.0066)
Treatment (t+0/Federal)	-0.0220***	-0.0252***	-0.0181
	(0.0063)	(0.0071)	(0.0141)
Treatment (t+0/State)	-0.0221***	-0.0234***	-0.0201
	(0.0065)	(0.0071)	(0.0140)
Treatment (t+1)	-0.0151***	-0.0288***	0.0049
	(0.0057)	(0.0076)	(0.0137)
Treatment (t+2)	-0.0231***	-0.0618***	0.0284**
	(0.0055)	(0.0066)	(0.0132)
Treatment (t+3)	-0.0193***	-0.0425***	0.0157
	(0.0063)	(0.0075)	(0.0142)
R^2	0.94967	0.91415	0.93792
Observations	12,150	12,150	12,150
Precinct FE	✓	✓	✓
Election-District FE	✓	✓	✓

Notes: The table presents event study results based on Equation 1 for postal turnout, in-person turnout, and total turnout controlling for additional variables. The reference election $(t-1)$ is the 2019 European Election. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.4 Clustering

In I baseline analysis, I cluster at the level of postal precincts, which is the level at which the treatment is applied. The district is the next higher administrative unit, with precinct administrators reporting to district offices. Concerns may arise that model errors are correlated within State constituencies. To address this, I increase the clustering level to State constituencies, as shown in Column 2 of Table A4. The standard errors are close to the baseline results when clustering at the county level. Additionally, two-way clustering—considering postal precincts, and district $\times$ elections results in slightly larger standard errors (Column 3) compared to clustering at the postal precinct level.

Table A4: Robustness: Standard Errors

Panel A: Total Turnout	(1)	(2)	(3)
Treatment (t-4)	0.0060	0.0060	0.0060
	(0.0062)	(0.0082)	(0.0088)
Treatment (t-3)	-0.0071	-0.0071	-0.0071
	(0.0049)	(0.0052)	(0.0065)
Treatment (t-2)	-0.0073	-0.0073	-0.0073
	(0.0060)	(0.0061)	(0.0072)
Treatment (t+0/Federal)	-0.0241***	-0.0241***	-0.0241***
	(0.0066)	(0.0054)	(0.0066)
Treatment (t+0/State)	-0.0243***	-0.0243***	-0.0243***
	(0.0067)	(0.0055)	(0.0067)
Treatment (t+1)	-0.0149**	-0.0149**	-0.0149*
	(0.0061)	(0.0058)	(0.0077)
Treatment (t+2)	-0.0185***	-0.0185***	-0.0185***
	(0.0056)	(0.0061)	(0.0060)
Treatment (t+3)	-0.0140*	-0.0140**	-0.0140
	(0.0078)	(0.0066)	(0.0096)
R^2	0.92949	0.92949	0.92949
Panel B: In-person Turnout	**(1)**	**(2)**	**(3)**
Treatment (t-4)	0.0133*	0.0133	0.0133
	(0.0073)	(0.0110)	(0.0129)
Treatment (t-3)	-0.0066	-0.0066	-0.0066
	(0.0050)	(0.0060)	(0.0079)
Treatment (t-2)	-0.0072	-0.0072	-0.0072
	(0.0074)	(0.0067)	(0.0097)
Treatment (t+0/Federal)	-0.0241***	-0.0241***	-0.0241**
	(0.0091)	(0.0072)	(0.0106)
Treatment (t+0/State)	-0.0222**	-0.0222***	-0.0222**
	(0.0090)	(0.0072)	(0.0104)
Treatment (t+1)	-0.0278***	-0.0278***	-0.0278**
	(0.0074)	(0.0074)	(0.0114)
Treatment (t+2)	-0.0619***	-0.0619***	-0.0619***
	(0.0063)	(0.0070)	(0.0086)
Treatment (t+3)	-0.0438***	-0.0438***	-0.0438***
	(0.0099)	(0.0078)	(0.0122)
R^2	0.86623	0.86623	0.86623
Panel C: Postal Turnout	**(1)**	**(2)**	**(3)**
Treatment (t-4)	-0.0089	-0.0089	-0.0089
	(0.0124)	(0.0139)	(0.0100)
Treatment (t-3)	-0.0062	-0.0062	-0.0062
	(0.0085)	(0.0119)	(0.0068)
Treatment (t-2)	-0.0044	-0.0044	-0.0044
	(0.0070)	(0.0114)	(0.0034)
Treatment (t+0/Federal)	-0.0191	-0.0191	-0.0191
	(0.0145)	(0.0121)	(0.0133)
Treatment (t+0/State)	-0.0214	-0.0214*	-0.0214
	(0.0145)	(0.0121)	(0.0132)
Treatment (t+1)	0.0084	0.0084	0.0084
	(0.0142)	(0.0120)	(0.0143)
Treatment (t+2)	0.0424***	0.0424***	0.0424***
	(0.0132)	(0.0121)	(0.0117)
Treatment (t+3)	0.0333**	0.0333***	0.0333**
	(0.0148)	(0.0127)	(0.0159)
R^2	0.93030	0.93030	0.93030
Observations	12,150	24,300	24,300
Standard-Errors	Precinct	Precinct-Election	Precinct & Election-District
Precinct FE	✓	✓	✓
Election-District FE	✓	✓	✓

Notes: The table presents event study results based on Equation 1 for the natural log of postal turnout, in-person turnout, and total turnout using different standard error calculation methods: Column (1) is the baseline and standard errors are clustered at the municipality level. In Column (2), standard errors are clustered at the State conctituency level. Column (3) uses two-way clustered standard errors at the level of postal precinct and District $\times$ elections. Standard errors are reported in parentheses. The reference election $(t-1)$ is the 2019 European Election. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.5 Matching on Observables

I conduct various matching procedures to ensure the comparability of the treatment and control group, namely propensity score matching, Entropy Balancing as proposed by Hainmueller (2012), and Mahalanobis distance matching. If applicable, I solely refer to values from the precinct-level structural data measured prior to the 2021 election. I re-evaluate the baseline model (Equation 1) using each matching method and present the findings in Appendix Table A5.

Local Matching First, I use a geographical local matching approach to ensure that the effect is not driven by outliers in the periphery, as postal precincts closer to each other may be more similar. I identify adjacent postal precincts within a district (within 10 meters of the boundaries).I match 295 treated units to 436 untreated units. All treatment effects are close to the baseline estimates and remain statistically significant.

Propensity Score Matching I conduct a propensity score matching procedure to estimate the likelihood of experiencing irregularities during the 2021 Federal elections. This propensity is calculated using a probit regression based on the following pre-treatment socioeconomic characteristics: population measured by the number of eligible voters, share of population aged 6-18, 18-25, 25-35, 35-35, 45-60, 60-70, and over 65, share of foreign population, share of EU citizens, share of females, share of citizens with migrant backgrounds, share of single households, and share of residents subject receiving social security. Additionally, I exact match on the district. All values are precinct-level structural data measured prior to the 2021 election. The matched sample is obtained using 1:1 nearest neighbor matching with replacement. In this process, 294 treated units are matched to 159 control units. The estimates derived from the matched sample confirm I original results for the effect of irregularities on total turnout with the exception of the 2023 State election, which loses significance.

Mahalanobis Matching I check the robustness of I findings by matching treated and untreated units based on the Mahalanobis distance. This method evaluates similarity by measuring the proximity of units' covariates in vector space. I calculate distances using the same pre-treatment covariates as previously mentioned and employ a 1:1 nearest neighbor matching with replacement to create the matched sample. In this process, 294 treated municipalities are matched to 162 control units, while 894 units are dropped. The results indicate that the treatment effects remain significant and are consistent with the outcomes of propensity score matching.

Entropy Balancing I implement the entropy balancing approach proposed by Hainmueller (2012). This method offers the advantage of not truncating the sample. Instead, it assigns a set of weights that balance the treatment and control groups across several moments of the covariate distributions. Specifically, I balance the means and variances of the pre-treatment characteristics listed above and use the resulting entropy weights in the event study regressions. The results consistently support I original findings across all specifications and outcomes.

Table A5: Matching

	Total Turnout			
	(1)	(2)	(3)	(4)
Treatment (t-4)	0.0056	0.0120	0.0086	0.0158**
	(0.0064)	(0.0074)	(0.0076)	(0.0071)
Treatment (t-3)	-0.0034	-0.0048	0.0022	-0.0061
	(0.0050)	(0.0063)	(0.0062)	(0.0060)
Treatment (t-2)	-0.0075	0.0001	0.0002	-0.0018
	(0.0062)	(0.0073)	(0.0075)	(0.0093)
Treatment (t+0/Federal)	-0.0240***	-0.0232***	-0.0229***	-0.0221***
	(0.0068)	(0.0082)	(0.0081)	(0.0078)
Treatment (t+0/State)	-0.0242***	-0.0248***	-0.0238***	-0.0212***
	(0.0069)	(0.0084)	(0.0083)	(0.0078)
Treatment (t+1)	-0.0131**	-0.0200***	-0.0132*	-0.0262***
	(0.0061)	(0.0074)	(0.0075)	(0.0097)
Treatment (t+2)	-0.0216***	-0.0192**	-0.0088	-0.0181**
	(0.0058)	(0.0075)	(0.0075)	(0.0089)
Treatment (t+3)	-0.0171**	-0.0107	-0.0017	-0.0168*
	(0.0081)	(0.0099)	(0.0101)	(0.0102)
R^2	0.93035	0.92680	0.92662	0.93399
Observations	6,561	4,086	4,104	11,034
Specification	Local Matching	Propensity Score	Mahalanobis	Entropy Bal.
Precinct FE	✓	✓	✓	✓
Election-District FE	✓	✓	✓	✓

Notes: The table presents event study results based on Equation 1 for the natural log of total turnout using different matching approaches. Column (1) uses nearest neighbour matching based on propensity score, and Column (2) based on Mahalanobis distance. Column (3) uses weights from Hainmueller (2012). The reference election $(t-1)$ is the 2019 European Election. Standard errors are clustered at the municipality level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

A.6 Spatial Sample Construction and Weighting

To assess the robustness of the main findings to different approaches for harmonizing spatial units over time, I re-estimate the core event study specification using two alternative constructions of the panel. Results are presented in Table A6.

Weighted Sample I weight each precinct-election observation by the maximum spatial overlap between a given election-year precinct and the 2021 postal precinct to which it was matched. This approach assigns greater weight to precincts with high spatial continuity across years and reduces influence from more fragmented or ambiguous matches. The results remain statistically and substantively consistent with the baseline: treatment effects are persistent and concentrated in in-person turnout, with partial substitution into postal voting.

Restricted Sample (99.9% Match) I restrict the sample to precinct-year pairs that share at least 99.9% of their population with a single 2021 postal precinct. This yields a smaller sample of 4,827 observations but ensures nearly perfect spatial comparability over time. Turnout effects remain negative and statistically significant for most elections post-treatment, particularly on the postal margin, where the estimated drop exceeds 8 percentage points immediately following the disruption. This suggests that treatment effects are not driven by imprecise geographic aggregation but reflect real behavioral responses.

Table A6: Spatial Sample Construction and Weighting

Panel A: Total Turnout	(1)	(2)
Treatment (t-4)	0.0015	
	(0.0068)	
Treatment (t-3)	-0.0073	
	(0.0058)	
Treatment (t-2)	-0.0082	
	(0.0070)	
Treatment (t+0/Federal)	-0.0249***	-0.0190**
	(0.0067)	(0.0086)
Treatment (t+0/State)	-0.0251***	-0.0192**
	(0.0067)	(0.0086)
Treatment (t+1)	-0.0156**	-0.0098
	(0.0061)	(0.0108)
Treatment (t+2)	-0.0215***	-0.0334**
	(0.0053)	(0.0162)
Treatment (t+3)	-0.0129	
	(0.0085)	
R^2	0.92902	0.95172
Panel B: In-person Turnout	**(1)**	**(2)**
Treatment (t-4)	0.0084	
	(0.0081)	
Treatment (t-3)	-0.0050	
	(0.0057)	
Treatment (t-2)	-0.0061	
	(0.0085)	
Treatment (t+0/Federal)	-0.0239**	0.0220**
	(0.0096)	(0.0107)
Treatment (t+0/State)	-0.0221**	0.0239**
	(0.0095)	(0.0108)
Treatment (t+1)	-0.0277***	0.0183
	(0.0076)	(0.0130)
Treatment (t+2)	-0.0627***	-0.0391**
	(0.0063)	(0.0190)
Treatment (t+3)	-0.0436***	
	(0.0111)	
R^2	0.86868	0.89018
Panel C: Postal Turnout	**(1)**	**(2)**
Treatment (t-4)	-0.0151	
	(0.0133)	
Treatment (t-3)	-0.0117	
	(0.0095)	
Treatment (t-2)	-0.0098	
	(0.0080)	
Treatment (t+0/Federal)	-0.0219	-0.0785***
	(0.0142)	(0.0151)
Treatment (t+0/State)	-0.0243*	-0.0808***
	(0.0143)	(0.0152)
Treatment (t+1)	0.0055	-0.0510***
	(0.0142)	(0.0150)
Treatment (t+2)	0.0372***	-0.0271*
	(0.0128)	(0.0149)
Treatment (t+3)	0.0361**	
	(0.0151)	
R^2	0.93007	0.95400
Observations	12,150	4,827
Specification	Weights	Constant
Precinct FE	✓	✓
Election-District FE	✓	✓

Notes: The table presents event study results based on Equation 1 for the natural log of total turnout using two alternative sample constructions. Column (1) weights each observation by the maximum share of population overlap between an election-year precinct and its matched 2021 postal precinct. Column (2) restricts the sample to precinct-year pairs with at least 99.9% population identity with a 2021 postal precinct. The reference election $(t-1)$ is the 2019 European Election. Standard errors are clustered at the municipality level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

B Data Sources and Harmonization

B.1 Main Administrative and Electoral Data

The core dataset is a harmonized precinct-level panel covering all Berlin state, federal, and European elections between 2014 and 2025. The unit of observation is the 2021 postal precinct (`BWB21`), which serves as the constant spatial frame across time. Election results are available at the precinct level (*Urnen-* and *Briefwahlbezirke*). Because precinct boundaries vary between elections, all historical data are harmonized to the 2021 delineations. For each election, I first aggregate results from in-person (*Urnenwahlbezirke*) and postal (*Briefwahlbezirke*) precincts to the `BWB` level, the smallest administrative unit consistently defined across elections. I then construct population-weighted spatial crosswalks linking each election's precinct geometries (2014, 2016, 2017, 2019, 2024, and 2025) to the 2021 configuration.

The crosswalks are based on 100-meter raster population data from the 2022 census (restricted to German citizens aged 18 and older), aggregated over the intersection of historical and 2021 precinct polygons. Within each source precinct, population shares are normalized to yield conversion weights that reallocate votes and socio-demographic variables to the 2021 boundaries. This approach primarily aggregates smaller historical precincts to larger 2021 units but proportionally divides overlapping areas where necessary. The resulting panel includes consistent precinct identifiers, official second-ballot results for all elections, harmonized socio-demographic indicators, and treatment flags for precincts affected by the 2021 election irregularities. Table A13 reports summary statistics.

A potential limitation of the harmonization is that precinct borders changed substantially between some elections, particularly prior to 2016. Fixing all observations to the 2021 delineations may thus introduce minor measurement errors if population density or voter composition differs within reallocated areas. Three factors mitigate this concern: (i) precincts are small, averaging roughly 1,500 eligible voters; (ii) the 100-meter census weighting ensures population-based rather than area-based reallocation; and (iii) robustness checks restricted to precincts without boundary or polling-place changes yield consistent estimates. Aligning all observations to the 2021 structure also facilitates comparison across pre- and post-rerun elections within a single, coherent spatial frame.

The dataset spans eight elections across six election years: European Parliament (2014, 2019, 2024), state (*Abgeordnetenhaus*) elections (2016, 2021, 2023), and federal (*Bundestag*) elections (2017, 2021). Precinct shapefiles are obtained from the Berlin Electoral Office.[5] Turnout and party vote shares are computed relative to eligible voters, with separate indicators for postal and in-person voting. I retain postal precincts as the analytical unit, as they provide complete coverage of both voting modes and cleanly map to treatment assignments defined by the 2021 rerun decision.

Socio-demographic controls come from Berlin's structural reports (*Strukturdaten*) for 2021, including population size, age composition, gender, migration background, and welfare depen-

[5] See, https://daten.berlin.de/datensaetze, last accessed on Oct 27, 2025.

dency rates, merged via unique precinct identifiers.[6] Administrative disruption data are derived from the Berlin State Court and Federal Constitutional Court rulings on the 2021 elections, which identify ballot shortages, waiting times, and other irregularities by polling station. These indicators are aggregated to the postal precinct (`BWB21`) level to define treatment and exposure intensities.

I quantify the severity of disruption in 2021 at the precinct level using official documentation from the State Constitutional Court (Verfassungsgerichtshof des Landes Berlin, 2022) and the Federal Constitutional Court (Bundesverfassungsgericht, 2023). The court record describes, for each annulled postal precinct, concrete administrative failures: polling stations that temporarily closed, stations that ran out of the correct ballots or distributed the wrong ballots, and queues that forced voters to wait well beyond the scheduled close of polls at 18:00 or to vote after 18:00 in violation of federal election law. I code these reports into a measure of waiting time (in hours) and aggregate them to the postal precinct, which allows us to estimate dose–response relationships rather than a simple treated/control contrast. Higher recorded waiting times in 2021 predict substantially larger and more persistent reductions in subsequent in-person turnout, as well as a gradual shift toward postal voting in later elections.

C Interpretation of the 2024 Federal Rerun.

The 431 precincts identified by the Federal Constitutional Court were subject to a court-ordered rerun of the Bundestag election in February 2024. While I retain these precincts in my analysis panel, I do not treat the rerun as a second intervention. The rerun occurred more than two years after the original disruption and involved only a subset of precincts, making causal inference impossible due to the lack of a control group. Contemporaneous reports and official statistics suggest extremely low turnout and limited campaign activity. The rerun was broadly perceived as a technical correction with no real political stakes. It is therefore unlikely to have reversed the behavioral or belief-based consequences of the 2021 disruption. I interpret the long-run turnout effects observed in subsequent regular elections as persistent scarring from the initial failure, not as artifacts of later rerun exposure.

C.1 Survey Data: GLES Rolling Cross Sections

To complement the administrative data, I draw on the GLES Rolling Cross-Section (RCS) surveys from 2017, 2021, and 2025. These surveys include repeated nationally representative samples of German citizens during each federal election campaign. I use only respondents aged 18 or older with valid information on residence (`State`), trust in election administration, and nonvoting timing.

The relevant trust item asks whether respondents agree with the statement: *"Die Bundestagswahl wurde von den zuständigen Behörden korrekt und fair durchgeführt"*. I construct three variables: (i) a binary indicator for those who "fully agree", (ii) a relaxed version includ-

[6] See, https://www.statistik-berlin-brandenburg.de/, last accessed on Oct 27, 2025.

ing “somewhat agree”, and (iii) an ordinal scale from -2 (“strongly disagree”) to $+2$ (“strongly agree”)).

For nonvoters, the timing of abstention is measured via post-election questions. I define two binary indicators: one for individuals who abstained late (on or shortly before election day), and one for those who cited procedural reasons for not voting. These outcomes allow testing for latent cost shocks, as outlined in the behavioral model.

D Supplementary Tables

Table A7: Pre-breakdown characteristics of treated and control precincts by election family

Panel A. Federal election baseline (2017)	Treated mean	Control mean	Diff (T - C)	SE	p-value
Turnout 2017	78.55	76.05	2.50	0.50	0.000
Share age 18–25	7.81	7.83	-0.02	0.16	0.877
Share welfare recipients (SGB II)	11.41	13.72	-2.31	0.61	0.000
Share migration background	13.62	13.90	-0.28	0.53	0.601
Panel B. State election baseline (2016)					
Turnout 2016	69.69	67.72	1.96	0.55	0.000
Share age 18–25	7.81	7.83	-0.02	0.16	0.877
Share welfare recipients (SGB II)	11.41	13.72	-2.31	0.61	0.000
Share migration background	13.62	13.90	-0.28	0.53	0.601
Panel C. European election baseline (2019)					
Turnout 2019	65.39	61.17	4.21	0.66	0.000
Share age 18–25	7.81	7.83	-0.02	0.16	0.877
Share welfare recipients (SGB II)	11.41	13.72	-2.31	0.61	0.000
Share migration background	13.62	13.90	-0.28	0.53	0.601

Notes: Each panel compares precincts later annulled by the Federal Constitutional Court ("treated") to other Berlin precincts ("control") in the last pre-breakdown election of the same type. Reported are mean turnout in that baseline election, and demographic composition: *share age 18–25*, *share of SGB II recipients* (means-tested welfare), and *share of residents with migration background*. "Diff (T - C)" is the treated-minus-control difference from an OLS regression of each row variable on the treatment indicator. Standard errors are in parentheses. Lower p-values indicate stronger statistical evidence of a difference. Turnout is in percent of eligible voters.

Table A8: Waiting Time

Panel A: Total Turnout	(1)	(2)
Treatment (t-4)	0.0060	0.0100
	(0.0062)	(0.0070)
Treatment (t-3)	-0.0071	-0.0035
	(0.0049)	(0.0053)
Treatment (t-2)	-0.0073	-0.0092
	(0.0060)	(0.0060)
Treatment (t+0/Federal)	-0.0241***	-0.0219***
	(0.0066)	(0.0061)
Treatment (t+0/State)	-0.0243***	-0.0223***
	(0.0067)	(0.0062)
Treatment (t+1)	-0.0149**	-0.0068
	(0.0061)	(0.0071)
Treatment (t+2)	-0.0185***	-0.0141*
	(0.0056)	(0.0072)
Treatment (t+3)	-0.0140*	-0.0232***
	(0.0078)	(0.0082)
R^2	0.92949	0.92941
Observations	12,150	10,854
Panel B: In-person Turnout	**(1)**	**(2)**
Treatment (t-4)	0.0133*	0.0169**
	(0.0073)	(0.0082)
Treatment (t-3)	-0.0066	-0.0019
	(0.0050)	(0.0053)
Treatment (t-2)	-0.0072	-0.0125
	(0.0074)	(0.0077)
Treatment (t+0/Federal)	-0.0241***	-0.0408***
	(0.0091)	(0.0095)
Treatment (t+0/State)	-0.0222**	-0.0398***
	(0.0090)	(0.0092)
Treatment (t+1)	-0.0278***	-0.0318***
	(0.0074)	(0.0091)
Treatment (t+2)	-0.0619***	-0.0466***
	(0.0063)	(0.0076)
Treatment (t+3)	-0.0438***	-0.0420***
	(0.0099)	(0.0106)
R^2	0.86623	0.86537
Panel C: Postal Turnout	**(1)**	**(2)**
Treatment (t-4)	-0.0089	-0.0058
	(0.0124)	(0.0131)
Treatment (t-3)	-0.0062	-0.0070
	(0.0085)	(0.0090)
Treatment (t-2)	-0.0044	-0.0024
	(0.0070)	(0.0057)
Treatment (t+0/Federal)	-0.0191	-0.0042
	(0.0145)	(0.0136)
Treatment (t+0/State)	-0.0214	-0.0055
	(0.0145)	(0.0136)
Treatment (t+1)	0.0084	0.0200
	(0.0142)	(0.0162)
Treatment (t+2)	0.0424***	0.0217
	(0.0132)	(0.0159)
Treatment (t+3)	0.0333**	-0.0004
	(0.0148)	(0.0162)
R^2	0.93030	0.92959
Observations	12,150	10,854
Specification	Baseline	Waiting Time in h
Precinct FE	✓	✓
Election-District FE	✓	✓

Notes: The table presents event study results based on Equation 1 for the natural log of postal turnout, in-person turnout, and total turnout. The reference election $(t-1)$ is the 2019 European Election. Treatment is defined as the maximum waiting time within a precinct as reported by the ruling of the *State Constitutional Court*. Standard errors are clustered at the postal precinct level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A9: Heterogeneity by Migrant Backgrounds

	Total Turnout		
	(1)	(2)	(3)
Treatment (t-4)	0.0061	0.0059	0.0051
	(0.0062)	(0.0062)	(0.0056)
Treatment (t-3)	-0.0068	-0.0068	-0.0061
	(0.0049)	(0.0048)	(0.0044)
Treatment (t-2)	-0.0088	-0.0072	-0.0055
	(0.0055)	(0.0060)	(0.0044)
Treatment (t+0/Federal)	-0.0214***	-0.0239***	-0.0228***
	(0.0064)	(0.0066)	(0.0063)
Treatment (t+0/State)	-0.0216***	-0.0242***	-0.0230***
	(0.0065)	(0.0067)	(0.0064)
Treatment (t+1)	-0.0107*	-0.0143**	-0.0155***
	(0.0055)	(0.0060)	(0.0059)
Treatment (t+2)	-0.0160***	-0.0185***	-0.0183***
	(0.0055)	(0.0056)	(0.0056)
Treatment (t+3)	-0.0133*	-0.0139*	-0.0114*
	(0.0074)	(0.0078)	(0.0062)
Treatment (t-4) $\times$ Z_i	0.0024	-0.0041	-0.0073
	(0.0070)	(0.0054)	(0.0067)
Treatment (t-3) $\times$ Z_i	-0.0033	-0.0013	0.0016
	(0.0058)	(0.0046)	(0.0059)
Treatment (t-2) $\times$ Z_i	-0.0004	-0.0025	0.0033
	(0.0057)	(0.0049)	(0.0053)
Treatment (t+0/Federal) $\times$ Z_i	-0.0133*	0.0054	-0.0085
	(0.0077)	(0.0061)	(0.0076)
Treatment (t+0/State) $\times$ Z_i	-0.0137*	0.0030	-0.0074
	(0.0078)	(0.0069)	(0.0079)
Treatment (t+1) $\times$ Z_i	-0.0114*	0.0112*	-0.0006
	(0.0063)	(0.0064)	(0.0073)
Treatment (t+2) $\times$ Z_i	-0.0074	0.0014	-0.0108*
	(0.0057)	(0.0048)	(0.0057)
Treatment (t+3) $\times$ Z_i	-0.0110	-0.0025	-0.0035
	(0.0084)	(0.0068)	(0.0077)
Covariate $Z_i	Migrant Backgrounds	Females	Welfare Recipients
R^2	0.93633	0.93023	0.94409
Observations	12,150	12,150	12,150
Precinct FE	✓	✓	✓
Election-District FE	✓	✓	✓

Notes: The table presents event study results based on Equation 2 for the natural log of total turnout. The reference election ($t-1$) is the 2019 European Election. The share of citizens with migrant backgrounds is held constant in 2019 and is scaled with a mean of 0 and a standard deviation of 1.Standard errors are clustered at the postal precinct level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A10: Heterogeneity by Age

	Total Turnout							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Treatment (t-4)	0.0063	0.0066	0.0063	0.0060	0.0043	0.0037	0.0046	0.0046
	(0.0062)	(0.0063)	(0.0061)	(0.0061)	(0.0063)	(0.0060)	(0.0063)	(0.0063)
Treatment (t-3)	-0.0083*	-0.0080	-0.0068	-0.0052	-0.0014	-0.0053	-0.0041	-0.0010
	(0.0050)	(0.0050)	(0.0050)	(0.0050)	(0.0045)	(0.0047)	(0.0048)	(0.0047)
Treatment (t-2)	-0.0087	-0.0083	-0.0083	-0.0066	-0.0026	-0.0038	-0.0050	-0.0031
	(0.0059)	(0.0062)	(0.0058)	(0.0061)	(0.0058)	(0.0055)	(0.0061)	(0.0061)
Treatment (t+0/Federal)	-0.0258***	-0.0260***	-0.0226***	-0.0228***	-0.0212***	-0.0223***	-0.0217***	-0.0214***
	(0.0066)	(0.0067)	(0.0065)	(0.0067)	(0.0066)	(0.0066)	(0.0066)	(0.0067)
Treatment (t+0/State)	-0.0258***	-0.0261***	-0.0228***	-0.0230***	-0.0214***	-0.0225***	-0.0218***	-0.0217***
	(0.0067)	(0.0068)	(0.0066)	(0.0068)	(0.0067)	(0.0067)	(0.0067)	(0.0068)
Treatment (t+1)	-0.0151**	-0.0163***	-0.0127**	-0.0116**	-0.0103*	-0.0167***	-0.0114*	-0.0091
	(0.0062)	(0.0061)	(0.0057)	(0.0059)	(0.0060)	(0.0061)	(0.0059)	(0.0059)
Treatment (t+2)	-0.0192***	-0.0190***	-0.0163***	-0.0176***	-0.0188***	-0.0189***	-0.0178***	-0.0184***
	(0.0056)	(0.0056)	(0.0055)	(0.0056)	(0.0056)	(0.0056)	(0.0055)	(0.0057)
Treatment (t+3)	-0.0163**	-0.0150*	-0.0137*	-0.0126	-0.0097	-0.0099	-0.0109	-0.0107
	(0.0077)	(0.0080)	(0.0076)	(0.0080)	(0.0078)	(0.0075)	(0.0079)	(0.0081)
Treatment (t-4) $\times$ Z_i	1.96×10^{-5}	0.0053	0.0044	0.0049	0.0083	0.0141***	-0.0118**	-0.0119**
	(0.0057)	(0.0058)	(0.0067)	(0.0051)	(0.0052)	(0.0050)	(0.0051)	(0.0055)
Treatment (t-3) $\times$ Z_i	-0.0109**	-0.0084*	-0.0033	-0.0075*	-0.0091**	-0.0156***	0.0143***	0.0105***
	(0.0046)	(0.0049)	(0.0050)	(0.0045)	(0.0039)	(0.0043)	(0.0043)	(0.0039)
Treatment (t-2) $\times$ Z_i	-0.0091	-0.0085	-0.0019	-0.0076	-0.0116***	-0.0136***	0.0142***	0.0126***
	(0.0057)	(0.0064)	(0.0047)	(0.0055)	(0.0042)	(0.0046)	(0.0052)	(0.0047)
Treatment (t+0/Federal) $\times$ Z_i	-0.0132**	-0.0166**	-0.0141**	-0.0132**	-0.0102**	-0.0084	0.0149***	0.0168***
	(0.0060)	(0.0067)	(0.0068)	(0.0056)	(0.0047)	(0.0055)	(0.0051)	(0.0056)
Treatment (t+0/State) $\times$ Z_i	-0.0112*	-0.0163**	-0.0140**	-0.0129**	-0.0096**	-0.0072	0.0154***	0.0154***
	(0.0061)	(0.0067)	(0.0071)	(0.0056)	(0.0048)	(0.0057)	(0.0052)	(0.0058)
Treatment (t+1) $\times$ Z_i	-0.0053	-0.0128**	-0.0064	-0.0116**	-0.0075	-0.0181***	0.0129**	0.0139**
	(0.0063)	(0.0062)	(0.0063)	(0.0052)	(0.0053)	(0.0063)	(0.0054)	(0.0057)
Treatment (t+2) $\times$ Z_i	-0.0067	-0.0049	-0.0121**	-0.0075	0.0012	0.0015	0.0019	0.0064
	(0.0052)	(0.0050)	(0.0049)	(0.0056)	(0.0049)	(0.0048)	(0.0050)	(0.0051)
Treatment (t+3) $\times$ Z_i	-0.0170**	-0.0088	-0.0134*	-0.0171**	-0.0124**	-0.0072	0.0216***	0.0176***
	(0.0080)	(0.0081)	(0.0072)	(0.0071)	(0.0059)	(0.0061)	(0.0067)	(0.0065)
Covariate $Z_i	Ages 0-6	Ages 6-18	Ages 18-25	Ages 25-35	Ages 35-45	Ages 45-60	Ages 60-70	Ages 70+
R^2	0.93119	0.93002	0.93293	0.93203	0.93210	0.93350	0.93134	0.93107
Observations	12,150	12,150	12,150	12,150	12,150	12,150	12,150	12,150
Precinct FE	✓	✓	✓	✓	✓	✓	✓	✓
Election-District FE	✓	✓	✓	✓	✓	✓	✓	✓

Notes: The table presents event study results based on Equation 2 for the natural log of total turnout. The reference election $(t-1)$ is the 2019 European Election. The share of citizens within the corresponding age group is held constant in 2019 and is scaled with a mean of 0 and a standard deviation of 1.Standard errors are clustered at the postal precinct level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A11: GLES Survey Results

A. " Election Process Correct and Fair"	(1)	(2)
Berlin 2017 (pre)	-0.0058	-0.0123***
	(0.0039)	(0.0032)
Berlin 2021 (post)	-0.0930***	-0.1275***
	(0.0035)	(0.0040)
Berlin 2025 (pre)	-0.0287***	-0.0075**
	(0.0034)	(0.0026)
Berlin 2025 (post)	-0.0175***	-0.0122***
	(0.0050)	(0.0031)
Specification	Strongly (Dis)agree	(Strongly) (Dis)agree
Observations	20,213	31,484
R^2	0.01346	0.01271
B. Timing of Voting Decision: Non-Vote	**(1)**	**(2)**
Berlin 2021 (post)	0.5281***	0.4950***
	(0.0360)	(0.0401)
Berlin 2025 (post)	-0.0077	0.1823**
	(0.0527)	(0.0621)
Specification	Prevented	Same Day | Prevented
Observations	510	510
R^2	0.04503	0.03429
Year FE	✓	✓
State FE	✓	✓

Notes: This table reports estimates from Equation 3 using pooled data from the GLES Rolling Cross-Section surveys conducted in 2017, 2021, and 2025. *Panel A* reports respondents' beliefs about electoral integrity, measured via agreement with the statement: *"Die Bundestagswahl wurde von den zuständigen Behörden korrekt und fair durchgeführt"*. The reference election (2021(*pre*)) is the *pre-election survey* of the 2021 Federal Election. Binary dependent variables equal 1 if respondents "fully agree" (Column 1), or "fully" or "somewhat agree" (Column 2); Column (3) codes responses from −2 ("strongly disagree") to +2 ("strongly agree"). *Panel B* examines the timing of abstention decisions among nonvoters. The reference election (2021(*pre*)) is the *pre-election survey* of the 2021 Federal Election. Column (1) codes as 1 those who decided not to vote in the final days before the election or reported being prevented from voting. Column (2) isolates respondents who cited inability to vote as the main reason. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A12: Election Results (Vote Share)

	LINKE	The Greens	SPD	FDP	CDU	AfD
	(1)	(2)	(3)	(4)	(5)	(6)
Treatment (t-4)	0.0876	-0.1042	-0.2446*	0.1520*	-0.3271**	0.3021*
	(0.1620)	(0.2046)	(0.1457)	(0.0835)	(0.1601)	(0.1832)
Treatment (t-3)	0.4236**	-0.4175**	0.2141	0.0230	-0.0321	-0.1042
	(0.1659)	(0.1925)	(0.1647)	(0.1273)	(0.1552)	(0.1125)
Treatment (t-2)	0.3675*	-0.7758***	0.1621	-0.1119	0.2166	-0.0363
	(0.2166)	(0.2530)	(0.1567)	(0.1330)	(0.1349)	(0.0768)
Treatment (t+0/Federal)	0.2277	0.3265	-0.0839	0.0079	-0.3843***	0.1245
	(0.1942)	(0.2293)	(0.1768)	(0.1273)	(0.1489)	(0.1064)
Treatment (t+0/State)	0.6126**	-0.1418	-0.2936	0.1507	-0.4197**	0.1059
	(0.2572)	(0.2019)	(0.2029)	(0.0992)	(0.1740)	(0.1026)
Treatment (t+1)	0.6545**	-0.1298	-0.0989	0.1728**	-0.8498***	0.0327
	(0.2654)	(0.2376)	(0.1491)	(0.0868)	(0.2903)	(0.1255)
Treatment (t+2)	0.3163	-0.1925	0.1630	0.0874	0.0550	-0.1902
	(0.2367)	(0.1881)	(0.1408)	(0.0895)	(0.1452)	(0.1484)
Treatment (t+3)	0.6977*	-0.4019*	0.0614	0.1501**	-0.1117	-0.2190
	(0.4215)	(0.2357)	(0.1452)	(0.0757)	(0.1680)	(0.1902)
R^2	0.95074	0.96863	0.91228	0.93247	0.96058	0.94452
Observations	12,150	12,150	12,150	12,150	12,150	12,150
Precinct FE	✓	✓	✓	✓	✓	✓
Election-District FE	✓	✓	✓	✓	✓	✓

Notes: The table presents event study results based on Equation 1 for the vote share of the political parties represented in the federal parliament *LINKE* (Left), *The Greens*, the *SPD* (Social Democrats), the *FDP* (Liberals), the *CDU* (Christian Conservatives), and the *AfD* (right-wing populist). The reference election $(t-1)$ is the 2019 European Election. Standard errors are clustered at the postal precinct level and reported in parentheses. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

Table A13: Summary Statistics

Panel A: Full Sample								
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Total Turnout	1,350	76.618	9.302	48.053	70.233	77.755	83.153	125.795
In-person Turnout	1,350	39.742	3.604	26.594	37.419	39.783	42.055	56.059
Postal Turnout	1,350	36.876	8.701	13.857	30.586	37.005	42.422	86.934
Eligible Voters	1,350	1,658.284	621.084	24.216	1,161.356	1,417.478	2,158.430	3,912.376
SGCwMB	1,350	2.475	3.108	0.024	0.733	1.494	3.088	35.697
Panel B: Treatment Group								
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Total Turnout	294	77.367	7.929	48.053	73.304	78.883	82.978	100.986
In-person Turnout	294	39.868	3.184	31.101	37.839	39.769	41.933	48.584
Postal Turnout	294	37.499	7.433	14.359	33.195	38.568	42.257	62.366
Eligible Voters	294	1,921.515	724.212	651.134	1,245.193	1,908.351	2,604.584	3,912.376
SGCwMB	294	2.833	3.203	0.026	0.878	1.979	3.583	27.875
Panel C: Control Group								
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Median	Pctl(75)	Max
Total Turnout	1,056	76.410	9.643	51.107	69.533	77.041	83.309	125.795
In-person Turnout	1,056	39.707	3.713	26.594	37.363	39.803	42.177	56.059
Postal Turnout	1,056	36.702	9.018	13.857	30.074	36.504	42.458	86.934
Eligible Voters	1,056	1,584.999	568.238	24.216	1,146.627	1,361.326	2,008.343	3,848.979
SGCwMB	1,056	2.375	3.076	0.024	0.687	1.378	2.886	35.697

Notes: The table presents summary statistics for the 2021 Federal election for the main sample, i.e. excluding those municipalities only affected by irregularities in the State election. It is split by treatment and control group for selected variables used in the analyses. *SGCwMB* is the share of German citizens older than 18 with migrant backgrounds.