

# Biased Electoral Polls

Henrik Sigstad\*

June 28, 2026

## Abstract

Do pre-election polls tilt toward the candidates who pay for them? In Brazil’s 2024 mayoral races, polls a candidate paid for overstate that candidate’s vote share by 7 percentage points compared to independent polls of the same candidate. The gap is not explained by any registered survey-design choice. Polling firms with substantial media work show less bias — consistent with reputational discipline. Surfacing past pollster bias to consumers of polls would let them discount biased polls and might also extend reputational discipline to more firms.

## 1 Introduction

Pre-election polls influence voters’ decisions through strategic voting and bandwagon effects.<sup>1</sup> Polls also signal viability to donors (Mutz, 1995) and to candidates considering whether to drop out or form alliances. A poll that overstates a candidate’s standing tilts voter and donor choices in that candidate’s favor.

Candidates thus have an incentive to influence the polls that measure their own support. Several documented cases suggest they do.<sup>2</sup> In Brazil’s 2022 presidential election, the polling

---

\*BI Norwegian Business School. Email: henrik.sigstad@bi.no.

<sup>1</sup>In multi-candidate races, voters use polls to identify viable candidates and avoid wasting their ballot on a non-contender (Granzier, Pons and Tricaud, 2023). Voters may also infer that widely-supported candidates are likely to be good, or simply prefer voting with the majority (McAllister and Studlar, 1991; Dahlgaard et al., 2015; Farjam, 2020; Araujo and Gatto, 2021).

<sup>2</sup>Candidate- and party-sponsored polls that diverge from the independent consensus are documented across diverse electoral systems. In Hungary, the polling firm Nézőpont Intézet, founded by a former Fidesz official and a documented recipient of large procurement contracts from Prime Minister Viktor Orbán’s government, published vote-intention polls systematically more favorable to Fidesz than independent firms ahead of the 2026 election (HVG360, 2025). In the United States, Trump 2024 campaign senior adviser Tony Fabrizio released a polling memo to donors in May 2024 claiming Trump led in all seven battleground states (Caputo, 2024). In Mexico, the polling firm Massive Caller, to which PAN paid more than 15 million pesos for surveys over 2018–2023 (Contralínea, 2024), showed PAN candidate Xóchitl Gálvez ahead of MORENA’s Claudia Sheinbaum in pre-election polls, while other major institutes had Sheinbaum leading by double-digit margins; she won by more than 30 percentage points.

firm Paraná Pesquisas received R\$ 2.7 million from President Bolsonaro’s party during the pre-campaign window. It reported the race tied between Bolsonaro and Lula while other major institutes had Lula ahead (Folha de S.Paulo, 2022). In the Brazilian state of Goiás in 2020, public prosecutors charged a single operator with selling 349 polls across 191 municipalities, each placing the paying candidate in first place.<sup>3</sup>

Whether candidates systematically pay for polls, and whether those polls are tilted in their favor, is a first-order question for the integrity of electoral information, not answered by the existing literature. Using data from the 2024 Brazilian municipal elections, this paper provides the first systematic evidence on this question. I measure how prevalent candidate sponsorship is, how much sponsored polls overstate their sponsor, where in the production chain the bias enters, and ask what could discipline it.

Brazilian electoral-poll regulation makes the question tractable. Every poll must be registered with the electoral court before release. The registration document names the contracting sponsor, declares the survey design, and is signed by a licensed statistician.

Candidate sponsorship is widespread. Of the 14,887 mayoral polls registered in 2024, 13.0% are commissioned by entities the registry links back to a specific candidate or party. Another 12.4% route through likely cover vehicles. The polling firm registers itself as sponsor on 31.7% of polls, which can also conceal candidate sponsorship. The 13.0% candidate-linked share is therefore a floor on the true incidence of candidate-sponsored polls.

I assemble a candidate-by-poll panel of 22,665 vote-intention observations across 2,669 municipalities and link each poll to its registered sponsor. On polls a candidate paid for, that candidate’s measured vote share averages +8.6 pp above their final result; on polls of the same races paid for by independent media or by the pollster itself, the same gap is +0.89 pp. The difference of +7.6 pp is statistically unchanged by candidate and race-by-week fixed effects. Polls commissioned by an opponent move the same candidate in the opposite direction by about −2.6 pp. The effect cannot be explained by candidates timing polls to when they are popular: independent polls of the same candidate fielded both the week before and the week after the sponsored poll show no contemporaneous movement in that candidate’s standing.

At what stage is the slant introduced? Not in any registered margin of the survey design. Sponsored polls do not oversample the strongholds of the candidate paying for the poll. They list the sponsor first *less* often than peer independents and describe their interviewer training and weighting procedures *more* thoroughly. Consistent with a formally unbiased survey design, I find no evidence that licensed statisticians refuse to sign sponsored polls. Standard digit-frequency forensics rule out crude post-fielding tampering. These analyses suggest that the slant is produced at stages the registration system does not see: operational departure from the declared sampling plan during fielding, or downstream sophisticated edits to the published numbers.

---

<sup>3</sup>Operação Leão de Neméia, MPE-GO 2020–22, against IPOP / Cidades e Negócios.

How can capture of polls be prevented? First, at least under current Brazilian laws, courts appear ineffective at addressing the bias: sponsored polls are no more likely to be sued for fraud than peer independents, suggesting the bias is induced at stages where courts cannot adjudicate. Second, several patterns are consistent with reputational incentives limiting the bias: firms that commission a substantial number of polls for media entities show essentially no bias on the polls they commission for candidates, and sponsored polls show less bias as the election approaches, when large deviations from the election result would be harder to explain.

The analysis suggests two policy responses: making the identity of sponsors and the past sponsor-bias of pollsters salient to consumers of polls, and making statistical evidence of bias in favor of sponsors a basis for legal accountability.

The closest prior literature studies how respondents react to a generic sponsor label in opt-in online surveys (Leeper and Thorson, 2019; Crabtree, Kern and Pietryka, 2020) and how sponsor identity affects stated exit-poll participation (Panagopoulos, 2016); the mechanism there (respondent perception of the sponsor) is distinct from the question studied here, where the respondent typically does not know the sponsor.

The paper also contributes to the literature on how voters and donors respond to polls (McAllister and Studlar, 1991; Dahlgard et al., 2015; Farjam, 2020; Granzier, Pons and Tricaud, 2023; Araujo and Gatto, 2021; Mutz, 1995) by documenting that these effects are large enough for candidates to seek to bias the polls measuring their own support.

The paper also contributes to the literature on Brazilian electoral polls—which has audited poll accuracy and decomposed error by sample design and voter-side dynamics (Gramacho, 2013; Meireles and Russo, 2022; Batista Pereira and Nunes, 2024)—by bringing in the identity of the sponsor as a factor.

The rest of the paper is organized as follows. Section 2 describes the Brazilian poll registration system. Section 3 describes the data. Section 4 sets out the research design. Section 5 presents the poll sponsor bias estimate. Section 6 asks at what production stage the slant is introduced. Section 7 discusses how to prevent capture of electoral polls. Section 8 concludes.

## 2 Setting

### 2.1 The pre-registration system

Brazilian law requires any organization releasing an electoral poll to register the poll with the electoral courts at least five business days before publication. The registration is required for any release, including releases internal to the sponsor’s campaign or to a small audience. Each registration document discloses the poll’s results along with the contracting sponsor, the paying entity, the funding source, and the declared cost. It also lists the sample size, the field pe-

riod, and the responsible statistician licensed by the regional statistical council. Three free-text methodology blocks describe the sampling plan, control system, and municipal coverage.<sup>4</sup>

## 2.2 Sponsor disclosure

Brazilian electoral law requires every registered poll to disclose its contracting sponsor’s taxpayer ID (CPF/CNPJ).<sup>5</sup> In principle the disclosure makes every sponsor administratively visible. In practice, polls funded by candidates can be routed through disguise mechanisms whose taxpayer ID does not link administratively back to the candidate. These include corporate shells, single-person sole-proprietor fronts, or registration as the polling firm’s own poll.

For instance, the polling firm IPOP-Cidades & Negócios in the state of Goiás was charged with selling 349 polls to candidates in 2020 (see Appendix A), each registered with the firm itself as the contracting sponsor. In 2024 the same firm had all 68 of its Goiás polls registered in the name of a private faculty in Goiânia (FacUnicamps), and no candidate appeared as sponsor. And in São Paulo, Publi. QC Pesquisas — which had self-contracted all 230 of its 2020 polls — routed 86% of its 2024 polls through a shell with R\$ 0 declared capital.

## 3 Data

The main data source is the TSE public-data portal for the 2024 elections (Tribunal Superior Eleitoral, 2024): the *pesquisa-eleitoral* poll registry, the candidate registry, the election results, and the donations filings. The poll registry records every mayoral poll registered in the 2024 cycle (14,887 polls), including the contracting sponsor’s taxpayer ID, the pollster, and the field period. Each registration carries the three free-text methodology blocks (see Section 2.1). I extract vote-intention shares for each scenario (the alternative candidate lists a single poll may ask about) from these reports with a large-language-model (LLM) pipeline (see Appendix C). The TSE candidate registry supplies the taxpayer ID, legal name, and ballot name (*nome de urna*) of every mayoral candidate. I use it to resolve sponsor taxpayer IDs to candidates. Candidate-level campaign donations come from the TSE donations filings. I use the Receita Federal firm registry (Receita Federal do Brasil, 2024) to identify media outlets (via economic-activity codes) and flag shell companies among the corporate sponsors (via share-capital fields).

### 3.1 Sample construction

I assemble a panel with one row per candidate per registered mayoral poll. I keep only the list-of-candidates scenario; where a single poll registers multiple such scenarios I keep the one

---

<sup>4</sup>Lei 9.504/97 Art. 33 enumerates the disclosed elements; TSE Resolução 23.600/2019 operationalizes them.

<sup>5</sup>Lei 9.504/97 Art. 33, inciso I.

with the most listed candidates,<sup>6</sup> so each poll contributes exactly one row per candidate.

I match poll-stated candidate names to the registered candidate using a token-overlap procedure.<sup>7</sup> I drop aggregate-name rows (blank, null, and “don’t know”) and rows whose listed-candidates scenario contains hypothetical names that cannot be matched. I keep only polls where every non-aggregate name matched a registered candidate.<sup>8</sup>

The resulting panel contains 22,665 candidate-poll observations across 7,908 candidates and 2,669 races. The four sponsor-recovery routes resolve to 450 candidate-poll rows in which the candidate’s own campaign sponsored the poll.

### 3.2 Sponsor classifier

I classify each registered poll’s contracting sponsor into one of six categories:

- *Candidate or party*: the sponsor resolves to a candidate or party through one of four recovery routes — the contracting taxpayer ID matches the candidate registry; the campaign-committee’s registered name matches the candidate; a party-owned taxpayer ID matches that party’s mayoral candidate in the municipality; or the free-text sponsor string itself names a party.
- *Media outlet*: identified through journalism economic-activity codes, restricted to firms with declared share capital of at least R\$10,000.<sup>9</sup>
- *Polling firm itself*: the contracting taxpayer ID is the pollster’s own.
- *Cover vehicle*: either a corporate shell (a firm taxpayer ID outside media and polling firms that commissions five or more polls in the cycle and whose registered name contains no media or polling keywords) or a sole-proprietor front (a single-person business that commissions polls).<sup>10</sup>
- *Residual*: uncoded.

The classifier yields the following shares of the 14,887 mayoral polls registered in 2024: 13.0% resolve to a specific candidate or party; 40.2% to media outlets; 31.7% to the polling firm itself;

---

<sup>6</sup>Ties on candidate-count are broken alphabetically on the registered sampling-plan description, for determinism.

<sup>7</sup>A poll-stated name is considered matched when it shares at least two tokens with either the candidate’s legal name or ballot name, or is a substring of either.

<sup>8</sup>The restriction drops  $\sim 20\%$  of candidate-poll rows; coefficients are essentially unchanged on the wider sample (Appendix D).

<sup>9</sup>The threshold excludes self-declared journalism firms whose registered capital is too small to support the operational base of journalistic activity (newsroom, equipment, payroll).

<sup>10</sup>Neither category has an obvious ordinary-business reason to commission polls at scale, consistent with vehicles that exist primarily to route registrations on behalf of an undisclosed sponsor.

12.4% to cover vehicles (7.7% corporate shells, 4.7% sole-proprietor fronts); 2.8% are uncoded. The 13.0% candidate-linked share is a lower bound on sponsor commissioning. Any poll funded by a candidate but routed through a cover vehicle or registered by the pollster enters the comparison group, not the treatment.

## 4 Research design

Do candidate-sponsored polls overstate the candidate who paid for them, and by how much?

To answer this question, I estimate the following regression:

$$y_{c,p} = \beta \text{Sponsored}_{c,p} + \gamma \text{Opp.Sponsored}_{c,p} + \mu_c + \rho_{r,t} + \varepsilon_{c,p}, \quad (1)$$

where  $y_{c,p}$  is the poll error,

$$y_{c,p} = (\text{poll share of candidate } c \text{ in poll } p) - (\text{final first-round vote share of } c),$$

expressed in percentage points;  $\text{Sponsored}_{c,p}$  equals 1 if candidate  $c$  paid for poll  $p$  and  $\text{Opp.Sponsored}_{c,p}$  equals 1 if another candidate in the same race paid for it;  $\mu_c$  is a candidate fixed effect; and  $\rho_{r,t}$  is a race-by-week fixed effect. Standard errors are clustered at the race level.

## 5 Results: bias in sponsored polls

Table 1 reports the estimates of running Equation 1. Column 1 shows that sponsored polls overstate the sponsor’s vote share by +7.6 pp compared to independent polls. Adding candidate fixed effects gives +6.8 pp (column 2); adding race-by-week fixed effects gives +6.9 pp ( $p = 0.000$ , column 3), comparing each candidate-sponsored poll to independent polls of the same candidate fielded in the same race and calendar week.<sup>11</sup>

Figure 1 plots the mean polling error by week around each self-sponsored poll. The sponsored-poll bin lies above the independent neighbors in the surrounding  $\pm 4$ -week window.

Polls sponsored by opponents show the candidate polling worse than independent polls.<sup>12</sup>

**Who commissions polls?** Within a race, candidates with higher final vote share are more likely to commission a self-sponsored poll: each 10-percentage-point increase in vote share is associated with about 1.5 pp higher commissioning probability. Self-sponsored polls also cluster

<sup>11</sup>Appendix D reports the standard robustness checks (drop largest sponsor, tighten name-matching, party-by-sponsor interactions, permutation, leave-one-out, and inference sensitivities); they leave the main coefficient within 0.5 pp.

<sup>12</sup>The magnitude is about a third of the self-sponsored coefficient, in part mechanically: poll shares sum to one, so an opponent-sponsor slant is distributed across multiple candidates’ numbers.

Table 1: Sponsor bias in registered 2024 Brazilian mayoral polls

|  | (1)                | (2)                | (3)                |
|--|--------------------|--------------------|--------------------|
| Self-sponsored $\hat{\beta}$ (pp)      | +7.63***<br>(0.66) | +6.77***<br>(0.82) | +6.86***<br>(1.33) |
| Opponent-sponsored $\hat{\gamma}$ (pp) | -0.93**<br>(0.47)  | -2.38***<br>(0.40) | -2.60***<br>(0.59) |
| Candidate FE                           |                    | ✓                  | ✓                  |
| Race $\times$ week FE                  |                    |                    | ✓                  |
| $N$                                    | 22,665             | 22,665             | 22,665             |

Each column is a separate regression of poll error (poll share – final vote share, in pp) on sponsored-by and opponent-sponsored indicators. Standard errors, clustered at the race (municipality) level, in parentheses. Additional intermediate specifications are reported in Appendix Table 5. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ . Sample: 22,665 candidate-by-poll observations across 7,908 candidates and 2,669 races.

closer to the election — median 13 days out versus 23 for independents. Appendix E reports the full set of candidate-level correlates.

## 6 How pollsters produce the slant

The most straightforward way to slant a poll is to oversample the sponsor’s strongholds — neighborhoods where the sponsor’s party did well in the prior cycle. For each poll I compute the sponsor’s party’s 2020 vote share in the neighborhoods (*bairros*) that poll declared sampling, using polling-station-level data.<sup>13</sup>

The results, presented in Table 2, return a null: sponsored polls sample neighborhoods with essentially the same sponsor-party vote share as matched independents.<sup>14</sup>

Another way to inflate the sponsor’s support is questionnaire priming: a sponsor could push respondents toward their own candidate by listing them first.<sup>15</sup> I do not find evidence of this. In fact, sponsored polls do this less often than matched independents.

Two further sample-composition margins fit the same pattern. Sponsored and independent polls use the registered-voter frame at near-identical rates (84.8% vs. 86.5%,  $p = 0.70$ ). And sponsored polls never use phone collection (0% vs. 9.8% for matched independents,  $p < 10^{-7}$ ); both groups overwhelmingly field in person.

A third way is operational discretion: a less-documented methodology leaves room to adjust

<sup>13</sup>See Appendix B for the matching procedure.

<sup>14</sup>A sharper version, using the sponsoring candidate’s own 2020 mayoral vote share (feasible on the 34 of 244 pairs where the candidate also ran for mayor in 2020 in the same municipality), is also null (+0.2 pp, cluster-bootstrap 95% CI [-0.5, +1.2] pp).

<sup>15</sup>Appendix B describes how this and the other rows of Table 2 are constructed.

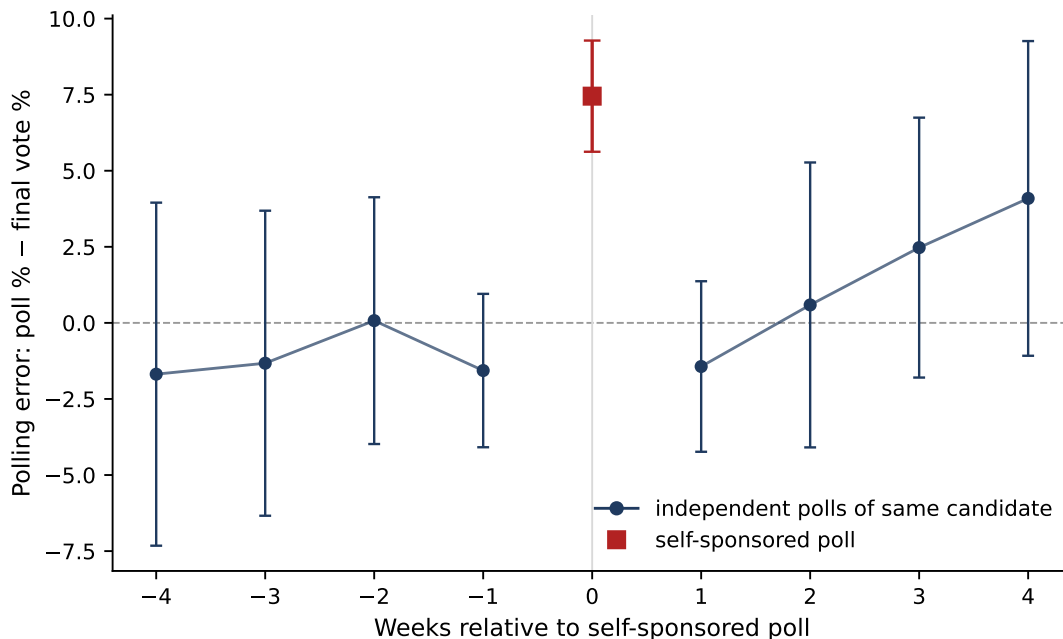


Figure 1: Mean polling error by week relative to the self-sponsored poll.

*Note:* Polling error is poll % – final vote % for independent polls of the same candidate in the same race, by week relative to the self-sponsored poll’s field-end date. Self-sponsored polls restricted to those with at least one independent neighbor in the  $\pm 4$ -week window ( $n = 116$ ). Error bars are 95% confidence intervals around the bin mean, clustered at the event level.

quotas or weights during fielding. I extract four documentation margins from the registered methodology PDFs: interviewer training, GPS geolocation, post-stratification weighting, and scenario rotation. In three of the four, sponsored polls describe these channels *more* thoroughly than peer independents — the opposite of the more-discretion prediction. The exception is scenario rotation, which sponsored polls document less often.

These tests suggest the slant is produced at stages the registration system does not see: operational departure from the declared design during fielding, or downstream edits to the published numbers.

In Appendix G, I test for crude per-candidate edits using standard digit-frequency forensics; the test rules them out. But sophisticated fabrication that preserves digit distributions remains consistent with the data.

## 7 How capture of electoral polls can be prevented

Two mechanisms might reach the bias: legal accountability and reputation. I document how each operates today and then ask how each could be sharpened.

Table 2: Survey-design margins: sponsored vs matched-independent polls

| Margin (share or mean of polls)                                 | Sponsored | Indep. | $p$               |
|---|-----------|--------|-------------------|
| <i>Substantive (sample composition)</i>                         |           |        |                   |
| Sponsor-party 2020 vote share in poll’s neighborhoods (mean)    | 23.4%     | 23.5%  | 0.40 <sup>a</sup> |
| Survey responses collected through phone calls (share of polls) | 0.0%      | 9.8%   | $< 10^{-7}$       |
| Population frame: registered voters                             | 84.8%     | 86.5%  | 0.70              |
| Sponsor’s candidate listed first in questionnaire               | 20.4%     | 30.1%  | 0.001             |
| <i>Documentation axis</i>                                       |           |        |                   |
| Interviewer training described in methodology PDF               | 84%       | 73%    | 0.002             |
| Interviews GPS-geolocated                                       | 18%       | 12%    | 0.058             |
| Post-stratification weighting explicitly described              | 60%       | 51%    | 0.056             |
| Scenario-rotation documented in questionnaire                   | 5.4%      | 26.1%  | $< 10^{-9}$       |

*Note:* Each row reports a per-poll share on matched race-week pairs (within  $\pm 14$  days,  $n=244$  each side), with the  $p$ -value from a sponsored-vs-independent Fisher exact (or two-proportion) test. <sup>a</sup>Paired-t on the 42 pairs with usable both-side neighborhood matches; the cluster-bootstrap 95% CI by sponsored-poll is  $[-0.46, +0.34]$  pp. Benjamini–Hochberg FDR correction on the  $m = 8$  displayed  $p$ -values: 4 survive at  $q < 0.05$  and 6 survive at  $q < 0.10$ .

## 7.1 Existing legal mechanisms

The Brazilian polling regime has two formal per-poll accountability mechanisms: legal sanctions for fraudulent dissemination of registered data, and the personal liability of the licensed statistician of record. This subsection covers legal sanctions; the statistician analysis is in Appendix F.

**Legal accountability.** I match each registered mayoral poll to the 2024 electoral-court case docket from Chin, Lambais and Sigstad (2024). I code a poll as the target of a fraud-allegation case if any 2024 case cites the poll’s registration protocol in its case text and the case’s subject classification (*assunto*) flags fraudulent dissemination or published-data irregularities.<sup>16</sup> Case-text protocol mapping covers 8,943 of the 14,887 registered polls. Table 3 reports a linear-probability regression of the fraud-allegation indicator on candidate-sponsored status.

Within race and fielding week, candidate-sponsored polls are no more likely to be targeted than peer independents — the coefficient is in fact negative and statistically significant (column 3:  $-0.039^{***}$ ,  $p < 0.01$ ).

The lawsuit-rate null suggests the documented slant falls in a gap in existing electoral law: no statutory provision directly targets subtle operational or post-collection manipulation of polls

<sup>16</sup>The two relevant *assunto* codes are *Divulgação de Pesquisa Eleitoral Fraudulenta* (fraudulent dissemination of an electoral poll) and *Irregularidades dos Dados Publicados em Pesquisas Eleitorais* (irregularities in published electoral-poll data).

Table 3: Probability of a 2024 electoral-court case targeting a registered poll

|                   | <i>Dependent variable: fraud-allegation case targeting poll</i> |                    |                      |
|-------------------|---|--------------------|----------------------|
|                   | (1)   | (2)                | (3)                  |
| Candidate sponsor | -0.011*<br>(0.006)  | -0.019*<br>(0.011) | -0.039***<br>(0.015) |
| Race FE           |   | ✓                  |                      |
| Race × week FE    |   |                    | ✓                    |
| N (polls)         | 8,943   | 8,943              | 8,943                |
| Mean of outcome   | 0.043   | 0.043              | 0.043                |

*Note:* Linear probability model of *fraud-allegation case targeting the poll* (1 if any 2024 fraud-allegation electoral-court case cites the poll’s registration protocol in its case text; mean 0.043) on a binary candidate-sponsor indicator (1 if the poll’s registered sponsor is a candidate or candidate committee; 0 if media, academic, or pollster-self). Universe is the 8,943 registered 2024 mayoral polls with case-text protocol coverage. Standard errors, clustered at the fixed-effect level, in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

produced under a disclosed methodology,<sup>17</sup> and statistical evidence of systematic bias across polls has never been tested in the courts.

## 7.2 Reputation

Polling firms depend on a record of accuracy to attract future customers: a firm publicly known to publish biased polls loses media contracts. Even candidates wanting biased polls have less incentive to commission a firm whose polls the market discounts as biased. Theoretically, reputation would discipline pollsters if (a) past bias is detectable to the consumers of polls and (b) the firm relies on clients for whom a reputation for honesty matters. If reputation operates, two patterns should hold:

1. firms whose portfolio carries more public reputational stake (a larger share of polls going to publicly identified media clients) show less bias;
2. bias shrinks where it would be more easily detected. This implies two specific patterns: within race, bias shrinks as the number of independent polls in the race grows (more comparators); and across time, bias shrinks as the election approaches (a shorter gap between the poll and the realized vote share makes it harder to attribute a divergence to genuine shifts in voter preference rather than to bias).

<sup>17</sup>Electoral-court jurisprudence interprets Lei 9.504/97 Art. 33 § 4<sup>o</sup> narrowly: the prosecuted misconduct must be the published numbers diverging from the registered methodology (Tribunal Superior Eleitoral, *Temas Seleccionados* → *Pesquisa eleitoral* → *Penalidade*, <https://temasseleccionados.tse.jus.br/temas-seleccionados/pesquisa-eleitoral/penalidade>). Reading the 2024 fraud-allegation docket confirms the pattern. Most filings allege procedural defects — missing methodological information, divergence between the registered and executed sampling plan, missing demographic strata, missing geographic delimitation. A smaller fraction allege bias in the published numbers, but when a case is granted with a bias allegation in the petition the court typically decides on procedural grounds.

I test these predictions by regressing the sponsor candidate’s signed percentage-point error per sponsored poll on covariates capturing each prediction: the media-share of the firm’s portfolio for the public-stake prediction (i); the log of the number of independent polls in the same race and the days from poll fielding to the election for the two detectability implications of (ii). I add the log of the firm’s total number of registered polls as a secondary reputational-stake measure: a firm registering more polls has more aggregate revenue to risk on a bias revelation. Standard errors are two-way clustered by race and firm.

Results are presented in Table 4. Two of the four coefficients reach the predicted direction with  $p < 0.05$ . A firm whose portfolio is entirely media-sponsored shows about 10 pp less per-poll bias than a firm with no media work (column 2,  $p = 0.04$ ), and each 100 additional days from the election predicts about +4.5 pp more bias (column 2,  $p = 0.004$ ). The within-race independent-polls coefficient and the firm-volume coefficient run in the predicted direction but neither is statistically significant. The pattern is consistent with reputation operating as a discipline on the per-poll bias: bias concentrates where the firm has the least reputational stake and where a biased poll would be hardest to detect.

Table 4: Per-poll sponsor bias on reputation-linked covariates.

|                               | (1)                | (2)                | (3)              |
|-------------------------------|--------------------|--------------------|------------------|
| Media-share of firm portfolio | -13.67**<br>(5.90) | -9.62**<br>(4.60)  | -0.70<br>(13.46) |
| log(firm volume)              | -0.62<br>(1.02)    | -0.66<br>(0.66)    | -1.44<br>(2.44)  |
| log(1 + indep. polls in race) | -1.38<br>(0.92)    | -1.11<br>(0.93)    |                  |
| Days to election (per 100)    | +1.78<br>(1.81)    | +4.49***<br>(1.56) | +1.79<br>(1.48)  |
| State fixed effects           |                    | ✓                  |                  |
| Race fixed effects            |                    |                    | ✓                |
| Observations                  | 520                | 520                | 520              |
| $R^2$                         | 0.061              | 0.223              | 0.939            |

**Notes.** One observation per sponsored poll (the sponsor candidate’s signed pp error,  $n = 520$  across 146 firms in 403 races). Firm-portfolio characteristics are computed over the firm’s full 2024 poll registry: media-share = the share of the firm’s polls with a publicly identified media client; log(firm volume) = log of the firm’s total registered polls. Race-level scrutiny is the log of 1+ the number of polls in the same race registered by an independent sponsor; this covariate is absorbed by race fixed effects in column 3. Days to election counts from the poll’s fielding date and is reported in units of 100 days. Standard errors are two-way clustered by race and firm. The race-fixed-effect specification (column 3) is identified off the small subset of races with multiple firms registered, leaving most firm-level coefficients underpowered. \*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$ .

### 7.3 Policy responses

I discuss two policy responses.

First, one could make bias more salient to consumers of polls. Regulators could require that the sponsor’s identity and the pollster’s past bias on candidate-sponsored polls be displayed alongside any poll results. Consumers of polls (voters and donors) could subtract the documented bias from each released poll, so biased polls would not move beliefs in the direction the sponsor intends. If voters discount sponsored headlines, candidates have less reason to pay for them, and pollsters less reason to slant.

The weakness of such a policy is that it gives candidates incentives to hide their sponsorship behind shell firms and other cover vehicles — a pattern already visible in the data (Appendix A documents one such case at scale). To bind, the scorecard would need to be coupled with stricter sponsor-identification rules at the registration step (requiring the actual source of funds for each poll to be disclosed and verifiable) and likely complemented by sponsor-agnostic measures for cases where the sponsor remains hidden, for example by publishing a pollster’s past accuracy in predicting electoral outcomes.<sup>18</sup>

Second, one might increase legal accountability for systematic pollster bias. One option is to increase what pollsters must disclose and punish survey designs that would inflate the sponsor’s stated support. But given that the formal survey designs in our data do not show evidence of bias (Section 6), more detailed methodology disclosure would likely relocate the slant to a margin the system does not yet cover rather than reduce it. A more promising route is liability based on cross-poll statistical patterns. Adjacent areas of law (employment-discrimination disparate-impact, securities fraud-on-the-market, antitrust cartel enforcement) routinely accept statistical evidence of a *class* of cases to establish liability where no individual case can be cleanly attributed. The same framework could in principle anchor a portfolio-level claim against a pollster whose polls systematically favor their sponsoring candidates; the Brazilian vehicle would be *Ação Civil Pública* (Lei 7.347/1985) or consumer-protection actions, though neither has been used against pollster bias patterns. Another possibility is to enforce personal liability for statisticians who sign off on biased polls.

## 8 Conclusion

Despite one of the world’s tightest pre-election poll registration regimes, candidates pay for at least 13% of registered mayoral polls in Brazil, and the polls they pay for overstate them by about 7 percentage points.

---

<sup>18</sup>Per-firm aggregate accuracy is uncorrelated with within-firm sponsor bias in the sample ( $r = -0.09$ ,  $p = 0.68$  across the 22 firms with within-firm estimates), so pure accuracy is not a sufficient signal of capture; a more sophisticated bias-detection metric would need to be developed.

This finding suggests that pre-election polls are, to a large extent, a strategic device, shaped by the candidates who commission them rather than measuring the voters they claim to survey. Sponsorship bias undermines the role polls play for voters and donors: candidates who can afford to commission them tilt the public information environment in their favor. Polls thus ought to be regulated carefully.

## References

Araujo, Víctor and Malu A. C. Gatto. 2021. “Casting Ballots When Knowing Results.” *British Journal of Political Science* 52(4):1709–1727.

Batista Pereira, Frederico and Felipe Nunes. 2024. “Pesquisas eleitorais e mudanças tardias na decisão do voto.” *Opinião Pública* 30:e3011.

Caputo, Marc A. 2024. “Tony Fabrizio feels good about Trump’s chances. And it’s ‘kinda weird’.” Accessed June 2026.  
<https://www.thebulwark.com/p/tony-fabrizio-feels-good-trump-polling>

Chin, Moya, Guilherme Lambais and Henrik Sigstad. 2024. “Constructing a Universe of Brazilian Electoral-Court Cases from DataJud.”. SSRN working paper.  
[https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5183038](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5183038)

Contralínea. 2024. “PAN ha pagado más de \$15 millones a Massive Caller, encuestadora que favorece a Gálvez.”. Accessed June 2026.  
<https://contralinea.com.mx/interno/semana/pan-ha-pagado-mas-de-15-millones-a-massive-caller->

Crabtree, Charles, Holger L. Kern and Matthew T. Pietryka. 2020. “Sponsorship Effects in Online Surveys.” *Political Behavior* 44(1):257–270.

Dahlgaard, J., J. H. Hansen, Kasper M. Hansen and Martin Larsen. 2015. “How do opinion polls affect voters? The effect of opinion polls on the Danes’ voting behavior and sympathy for parties.” *Politica* 47(1).

Farjam, Mike. 2020. “The Bandwagon Effect in an Online Voting Experiment With Real Political Organizations.” *International Journal of Public Opinion Research* 33(2):412–421.

Folha de S.Paulo. 2022. “Paraná Pesquisas recebeu R\$ 2,7 milhões de partido de Bolsonaro na pré-campanha.”. Accessed June 2026. Indexed in references/news/stories.csv.  
<https://www1.folha.uol.com.br/poder/2022/09/parana-pesquisas-recebeu-r-27-milhoes-de-partido-shtml>

Gramacho, W. 2013. “À margem das margens? A precisão das pesquisas pré-eleitorais brasileiras em 2010.”.

<https://www.scielo.br/j/op/a/Qg95LjqPvrBJf37kCZYrjnM/?lang=pt&format=pdf>

Granzier, Riako, Vincent Pons and Clémence Tricaud. 2023. “Coordination and Bandwagon Effects: How Past Rankings Shape the Behavior of Voters and Candidates.” *American Economic Journal: Applied Economics* 15(4):177–217.

HVG360. 2025. “Nézőpont: profitorientáltan alakították át a kormányhoz közeli közvéleménykutatót.”. Accessed June 2026.

[https://hvg.hu/360/20250724\\_nezopont-intezet-mraz-agoston-kozvelemenykutatas-hvg](https://hvg.hu/360/20250724_nezopont-intezet-mraz-agoston-kozvelemenykutatas-hvg)

Jornal Opção. 2024. “Veja lista de 40 cidades que instituto é suspeito de fraudar sondagem eleitoral.”. Accessed June 2026.

<https://www.jornalopcao.com.br/ultimas-noticias/veja-lista-de-40-cidades-que-instituto-e-sus>

Leeper, Thomas J. and Emily Thorson. 2019. “Should We Worry About Sponsorship-Induced Bias in Online Political Science Surveys?” *Journal of Experimental Political Science* 7(3):209–217.

McAllister, Ian and Donley T. Studlar. 1991. “Bandwagon, Underdog, or Projection? Opinion Polls and Electoral Choice in Britain, 1979–1987.” *The Journal of Politics* 53(3):720–741.

Meireles, Fernando and Guilherme Russo. 2022. “Pesquisas eleitorais no Brasil: tendências e desempenho.” *Estudos Avançados* .

<https://www.scielo.br/j/ea/a/HpJD5WZ485T56SmKkkwSbRz/?lang=pt&format=pdf>

Ministério Público do Estado de Goiás. 2020. “Leão de Neméia: MPE cumpre mandados de busca e apreensão na casa de dono de instituto de pesquisa.”. Accessed June 2026. Operation pursued by MPE-GO 2020–2022.

<https://www.mpgo.mp.br/portal/noticia/leao-de-nemeia-mpe-cumpre-mandados-de-busca-e-apreensa>

Mutz, Diana C. 1995. “Effects of Horse-Race Coverage on Campaign Coffers: Strategic Contributing in Presidential Primaries.” *The Journal of Politics* 57(4):1015–1042.

Panagopoulos, Costas. 2016. “Exit Poll Sponsorship and Response Intentions.” *Journal of Politics and Law* 9(4):72–80.

Receita Federal do Brasil. 2024. “Cadastro Nacional da Pessoa Jurídica (CNPJ).” Public firm registry maintained by Receita Federal. 2024 snapshot.

<https://dadosabertos.rfb.gov.br/CNPJ/>

Tribunal Superior Eleitoral. 2024. “Repositório de Dados Eleitorais.”. Public-data portal for Brazilian elections. Accessed June 2026.

<https://dadosabertos.tse.jus.br/>

## A Poll-selling and cover vehicles: the Goiás case

This appendix uses the Goiás polling firm IPOP-Cidades & Negócios to illustrate two phenomena documented in the body: (i) polls being explicitly sold to candidates and (ii) cover-vehicle substitution that hides the candidate-sponsor link from the registry.

IPOP-Cidades & Negócios self-contracted 357 mayoral polls across 192 municipalities in 2020, naming itself as contracting sponsor in every case. *Operação Leão de Neméia* (MPE-GO, 2020–2022) prosecuted the firm for 349 of these polls as fraudulent. The pollster charged candidates approximately R\$ 6,000 per poll to be placed in first position (Ministério Público do Estado de Goiás, 2020). The case is direct evidence of polls being sold as products to candidates.

The firm continued operating in 2024, but the registration pattern changed: its 68 mayoral polls in Goiás that cycle were all formally contracted by a private faculty in Goiânia (FacUnicamps), with no candidate appearing as contracting sponsor. The arrangement was contested in a 2024 *representação* (formal complaint) at the Goiás electoral court, where a candidate’s attorney questioned who had actually paid; FacUnicamps did not respond to press inquiries about its role in the polls (Jornal Opção, 2024).

The use of cover vehicles to hide the real contractor of polls is not local to Goiás. The 14 highest-volume sponsors that the registration data does not link to a candidate, party, media outlet, or pollster firm together account for 668 polls across 15 states. For instance, in São Paulo, VS Publicidade LTDA (R\$ 0 declared social capital, no public web presence) registered as contracting sponsor for 254 polls in 2024, 218 (86%) of which were produced by Publi. QC Pesquisas & Editoração — a pollster that had self-contracted all 230 of its own mayoral polls in 2020.

## B Construction of survey-design margin variables

Table 2 reports per-poll values on eight survey-design margins. Each is extracted from the poll’s official TSE registration document via a structured-output LLM extraction pipeline.

For each margin below, the extractor name, the verbatim prompt fragment that produces it, and the schema field whose value defines the margin are reproduced inline.

- *Sponsor’s party 2020 vote share in poll’s neighborhoods*: per-poll value is the bairro-weighted mean of the sponsoring candidate’s party’s 2020 mayoral vote share at the polling-station (*seção*) level, weighted by each neighborhood’s share of the municipality’s voter population.

Extractor: `poll_bairro_detail` (over the *bairro/município detail* PDF). Prompt fragment: “You extract structured coverage information ... listing the bairros / localidades

- / setores censitários where the poll was actually conducted.*” Schema field: `bairros: list[BairroEntry]` with each entry carrying a `bairro_name` (string) and `n_interviews` (int). The per-poll value is constructed by joining each `bairro_name` to the TSE polling-station registry and aggregating 2020 vote shares weighted by neighborhood population.
- *Survey responses collected through phone calls:* binary indicator for whether the methodology PDF’s declared collection mode is *phone* (as opposed to in-person, online, or mixed).  
Extractor: `poll_operations`. Prompt fragment: “*‘entrevistas telefônicas’, ‘CATI’ → mode=phone.*”  
Schema field: `mode: "in_person" | "phone" | "online" | "mixed" | "not_specified"`.  
Margin = 1 if `mode = "phone"`.
  - *Population frame: registered voters:* binary indicator for whether the methodology PDF’s declared target population is the standard registered-voter universe (as opposed to a hybrid frame combining the voter roll with Census data, or a Census-only frame).  
Extractor: `poll_sampling`. Prompt fragment: “*‘eleitorado’ + TSE source → tse\_eligible. if both census and TSE are cited and used at different stages → mixed.*” Schema field: `population_reference: "census_2022_residents" | "tse_eligible" | "mixed" | "not_specified"`.  
Margin = 1 if `population_reference = "tse_eligible"`.
  - *Sponsor’s candidate listed first in questionnaire:* from the questionnaire’s candidate-name ordering field, binary indicator for whether the sponsoring candidate appears in position 1.  
Extractor: `poll_questionario`. Prompt fragment: “*‘candidates\_listed: up to 15 entries per scenario, in the exact order they appear in the PDF. Preserve original capitalization.’*”  
Schema field: `candidates_listed: list[str]` per scenario. Margin = 1 if the sponsor’s candidate name matches `candidates_listed[0]` of the first estimulada scenario.
  - *Interviewer training described in methodology PDF:* binary indicator for whether the methodology PDF describes a training program for fieldwork interviewers.  
Extractor: `poll_operations`. Prompt fragment: “*‘treinados’, ‘treinamento’, ‘instruções específicas’ → interviewer\_training\_described=true.*” Schema field: `interviewer_training_described: bool`. Margin is the field’s value directly.
  - *Interviews GPS-geolocated:* binary indicator for whether the methodology PDF declares GPS-tagging of each interview.  
Extractor: `poll_operations`. Prompt fragment: “*‘100% digital’ + ‘geolocalizado’ + face-to-face → geolocated=true.*” Schema field: `geolocated: bool`. Margin is the field’s value directly.
  - *Post-stratification weighting explicitly described:* from a separate weighting extractor, binary indicator for whether the methodology PDF explicitly describes post-fielding weighting toward the target population.

Extractor: `poll_weighting` (Portuguese-language prompt). Prompt fragment: “*Marque post\_stratification\_explicit=true somente quando o texto AFIRMAR explicitamente que os pesos normalizam a amostra para a distribuição da população/alvo.*” Schema field: `post_stratification_explicit: bool`. Margin is the field’s value directly.

- *Scenario-rotation documented in questionnaire*: binary indicator for whether the questionnaire describes randomized rotation of vote-intention scenarios or candidate names within each scenario.

Extractor: `poll_questionario`. Prompt fragment: “*Set name\_rotation\_documented=true ONLY when the PDF prints an explicit instruction to rotate. Most PDFs have a fixed printed order — that is NOT rotation.*” Schema field: `name_rotation_documented: bool`. Margin is the field’s value directly.

Within-pair tests use Fisher’s exact (or two-proportion, where applicable) on the per-poll indicator. The stronghold row uses a paired-t on the per-pair difference in neighborhood-weighted-mean shares, with the cluster-bootstrap standard error by sponsored poll.

## C LLM extraction of vote-intention shares

The per-candidate vote-intention shares used as the outcome are extracted from each poll’s TSE *relatório* PDF (the methodology and results report that accompanies the registration) by a structured-output LLM pipeline. PDF text is obtained by `pdftotext -layout`; image-only PDFs (where `pdftotext` returns near-zero characters) are skipped. The remaining text is sent to GPT-4o-mini together with a narrow Pydantic schema that the model is required to populate via OpenAI’s Structured Outputs (constrained decoding), so the output cannot drift on field names or types.

**System prompt and schema.** The complete extractor prompt and JSON schema are reproduced verbatim below.

```
You extract per-candidate vote intentions from Brazilian TSE
relatório PDFs (text extracted by pdftotext).
```

```
Each PDF reports ONE registered poll. The PDF may reference
earlier waves --- extract ONLY the current/latest poll’s results,
not the historical comparison values.
```

Conventions:

- "Estimulado" = stimulated (names read).
- "Espontâneo" = spontaneous (open-ended).
- "Votos válidos" = valid votes (excludes Brancos/Nulos/Indecisos).
- Some institutes report rejection ("rejeição"), government evaluation ("avaliação"), and second-round simulations alongside vote intention --- include them as separate scenarios.
- For aggregate rows like "Branco/Nulo", "Não sabe", emit a candidate entry with party=null and a descriptive candidate\_name.
- TSE registration number is on every PDF, format "XX-NNNNN/YYYY".
- We DO NOT need methodology, dates, sample size, institute, contracting party --- those join from a separate CSV. Focus on the vote intention numbers.

If the text is too garbled to extract anything, return an empty scenarios list and explain in extraction\_notes.

Return ONLY a JSON object with EXACTLY the following structure (no other top-level keys, no other keys inside scenarios or candidates):

```
{
  "tse_protocol": "XX-NNNNN/YYYY",
  "scenarios": [
    {
      "scenario_type":
        "espontaneo" | "estimulado" | "votos_validos"
        | "rejeicao" | "avaliacao_governo"
        | "segundo_turno_simulacao" | "outro",
      "scenario_label": "the exact label used in the PDF",
      "candidates": [
        {
          "candidate_name": "display name or aggregate label",
          "party": "PL" | "PT" | ... | null,
          "percent": 0.0
        }
      ]
    }
  ]
}
```

```

    }
  ],
  "extraction_notes": ""
}

```

scenario\_type must be one of the seven lowercase ASCII labels listed (espontaneo without tilde, rejeicao without cedilla, etc.). percent is a number 0--100. party is null when absent. Do not invent keys: the field names are tse\_protocol, scenarios, scenario\_type, scenario\_label, candidates, candidate\_name, party, percent, extraction\_notes.

The main analysis uses only the *estimulado* scenarios. Institutional metadata (institute, dates, sample size, municipality, declared methodology) joins from the TSE registration CSV by registration protocol and is never re-extracted.

**Validation.** The Pydantic schema is enforced both server-side (via Structured Outputs) and client-side after return; entries that fail validation are dropped from the extraction cache. Downstream sanity checks confirm that the per-poll vote-intention denominators sum near 100 for the estimulado scenarios used in the analysis; the small share of polls with denominators far from 100 is excluded by the name-match restriction in Section 4.

**Limitations.** Image-only PDFs are skipped rather than OCR'd; an OCR fallback is left for a future revision. The extracted candidate strings are matched to the TSE candidate registry by a token-overlap procedure (Section 4), which is the source of any remaining name-match noise.

## D Robustness

Table 5 reports three variant specifications: candidate FE without race-time FE; a clean-comparator restriction that excludes opponent-sponsored polls, isolating the sponsored-vs-truly-independent contrast; and race  $\times$  month FE as a coarser-time analog of the main race  $\times$  week. The within-candidate coefficient sits in a [+6.8, +7.8] pp band across these variants.

Table 6 reports sample and name-matching sensitivities: dropping the largest single sponsor (4 polls) and tightening the candidate-name match to require at least three (or four) matching tokens leave both main coefficients within a percentage point of the baseline. Table 7 reports inference sensitivities: two-way municipality  $\times$  pollster and candidate-level clustering leave the within-candidate  $p$ -value far below 0.01, and wild-cluster restricted bootstrap  $p$ -values (2,000 Rademacher draws at the municipality level) round to below 0.001 for both main specifications.

Table 5: Intermediate specifications (appendix)

|  | Cand FE only       | Clean comparator   | Race $\times$ month FE |
|--|--------------------|--------------------|------------------------|
| Self-sponsored $\hat{\beta}$ (pp)      | +6.77***<br>(0.82) | +6.57***<br>(1.33) | +7.84***<br>(1.57)     |
| Opponent-sponsored $\hat{\gamma}$ (pp) | -2.38***<br>(0.40) | -                  | -                      |
| Candidate FE                           | ✓                  | ✓                  | ✓                      |
| Pollster FE                            |                    | ✓                  |                        |
| Race $\times$ month FE                 |                    |                    | ✓                      |
| $N$                                    | 22,665             | 16,125             | 16,125                 |

Intermediate specifications between the headline columns in Table 1. *Cand FE only* reports the within-candidate coefficient without pollster fixed effects, showing that adding pollster FE moves the coefficient very little. *Clean comparator* restricts the comparator pool to independent-media and pollster-self polls without imposing a race-time FE. *Race  $\times$  month FE* is the race-by-month analog of the headline race-by-week specification. Standard errors, clustered at the race (municipality) level, in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Figure 2 reports the within-candidate  $\hat{\beta}$  under leave-one-pollster-out (top 20 firms) and leave-one-UF-out (26 states) refits. The pollster refits range over [+6.9, +7.3] pp and the state refits over [+6.6, +7.3] pp; no single firm or state drives the estimate. A row-level sponsor-label permutation under candidate FE produces a null centered on zero (sd 0.59 pp, max  $|\hat{\beta}|$  over 500 draws = 1.8 pp) against the observed +6.8 pp,  $p < 0.002$ .

Re-estimating the within-candidate specification with five self-sponsored-by-party interactions (PSD, PL, PP, MDB, UNIÃO, with all other parties pooled into a baseline) yields a joint Wald test of  $\chi^2(5) = 2.3$  ( $p = 0.805$ ). The implied per-party coefficients range from +3.9 pp (PP) to +11 pp (UNIÃO); every party-specific confidence interval overlaps the pooled +6.8 pp.<sup>19</sup>

## E Selection patterns: who commissions polls?

The within-candidate fixed-effects design identifies the sponsor bias on the cohort of candidates who do commission polls. This appendix documents who those candidates are.

**Candidate-level correlates.** Table 8 reports a linear-probability regression of the binary *ever self-sponsored* indicator (one if the candidate paid for at least one of their own polls) on candidate-level covariates, with race fixed effects and standard errors clustered at the municipality. Of the  $N = 8,308$  candidate-races in the poll-covered sample, 5.2% commission at least one poll. Within

<sup>19</sup>Scenario-pick robustness: 93% of estimulado-only mayoral protocols have a single scenario and the canonical-pick choice is degenerate; on the remaining 616 multi-scenario protocols, the within-candidate median vote-percent spread across scenarios is 3.2 pp, so the aggregate scenario-pick effect on  $\hat{\beta}$  is bounded at well under 1 pp.

Table 6: Sample and name-matching sensitivities of the main coefficients

|  | Within-cand. $\hat{\beta}$ (pp) | Race $\times$ week $\hat{\beta}$ (pp) |
|--|---------------------------------|---------------------------------------|
| Baseline (Table 1 cols. 2–3)                       | +7.04***<br>(1.13)              | +7.69**<br>(3.52)                     |
| Drop the largest single sponsor                    | +7.04***<br>(1.13)              | +7.69**<br>(3.52)                     |
| Require $\geq 3$ matching tokens in candidate name | +6.80***<br>(1.12)              | +7.65**<br>(3.56)                     |
| Require $\geq 4$ matching tokens in candidate name | +6.94***<br>(1.12)              | +7.65**<br>(3.55)                     |

Each row refits the two headline specifications on a perturbed sample. Within-candidate column reports  $\hat{\beta}$  on Sponsored with candidate fixed effects only; race  $\times$  week column adds race  $\times$  week fixed effects on the strict clean-comparator subset. Standard errors, clustered at the race (municipality) level, in parentheses. \* $p < 0.10$ , \*\* $p < 0.05$ , \*\*\* $p < 0.01$ .

Table 7: Inference sensitivities for the within-candidate and race  $\times$  week main specifications

|  | Within-cand. | Race $\times$ week |
|--|--------------|--------------------|
| $\hat{\beta}$ (pp)                         | +6.77        | +6.86              |
| SE, municipality cluster (baseline)        | 0.82         | 1.33               |
| SE, municipality $\times$ pollster two-way | 1.01         | —                  |
| SE, candidate cluster                      | 0.84         | —                  |
| Wild-cluster restricted $p$ (B=2,000)      | <0.001       | <0.001             |

Cluster-robust standard errors under alternative clustering schemes for the within-candidate specification. Wild-cluster restricted bootstrap  $p$ -values use 2,000 Rademacher draws clustered at the municipality level. Dashes indicate the variant is not applicable to that specification.

a race, the probability of commissioning a poll rises with final vote share: a 10-percentage-point higher final share is associated with about 1.5 percentage points higher commissioning probability (s.e. 0.5). The quadratic term is null within race (a linear approximation suffices in the relevant range), and log campaign donations does not predict the commissioning decision.

## F Do statisticians refuse to sign biased polls?

Beyond the declared design, the registration record makes one further production stage visible: the responsible statistician of record. Every poll registration carries the name and license number of a CONRE-registered statistician who signs the declared methodology under personal legal liability.<sup>20</sup>

<sup>20</sup>Lei 9.504/97 Art. 35; Lei 4.739/1965 Art. 11.

Jackknife stability of the spec-2 sponsor coefficient

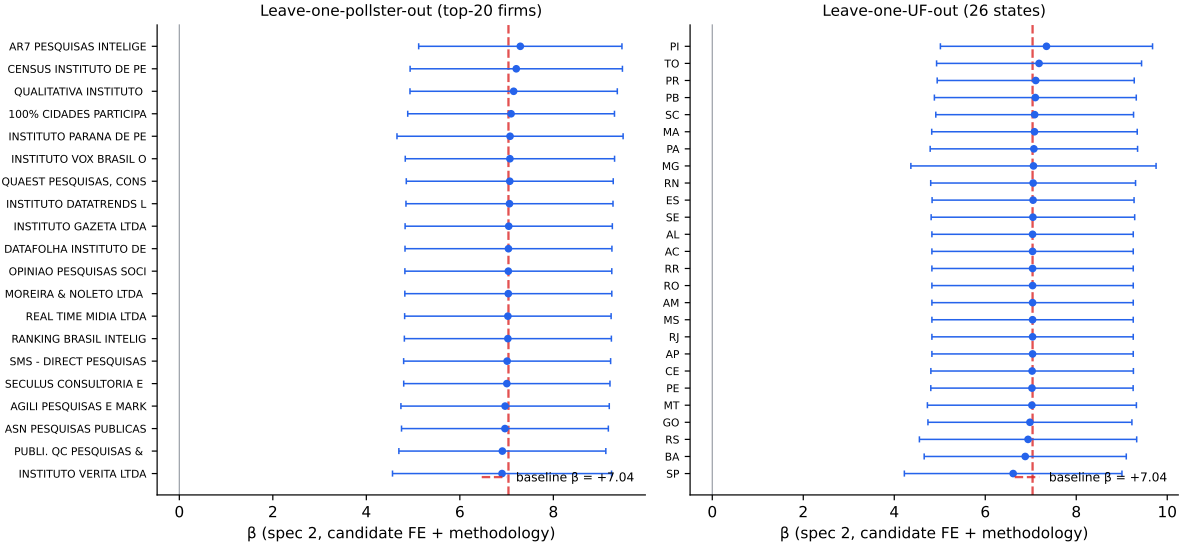


Figure 2: Leave-one-pollster-out (top 20 firms) and leave-one-UF-out (26 states) refits of the within-candidate  $\hat{\beta}$ . Dashed line: baseline estimate.

If statisticians refuse to sign biased polls, sponsored polls should concentrate on a subset of signatories within each firm. I test this on the 19 firms with at least two signatories, at least two sponsored polls, and at least two unsponsored polls: for each firm I compute the variance of the sponsored-poll share across its signatories, aggregated across the 19 firms. The observed spread is 0.140; under a within-firm permutation null (reassigning sponsored status to polls within each firm at random, recomputing the spread,  $B = 500$ ), the null mean is 0.157 (sd 0.026), giving  $p = 0.73$ . Statisticians do not appear to refuse biased polls—suggesting that the slant is produced at stages the signing statistician cannot observe.

## G Digit-frequency forensics

I run a two-proportion  $z$ -test on the frequency of round-number reporting between the sponsor's own candidate ( $n_A = 627$  rows) and other candidates in the same sponsored polls ( $n_B = 1,182$  rows). The result is null:  $z = +0.26$ ,  $p = 0.792$ . The sponsor's reported number is digit-indistinguishable from the other candidates' numbers in the same poll, so crude per-candidate post-fielding tampering is not detected. Sophisticated edits that preserve digit distributions cannot be ruled out.

Table 8: Candidate-level selection: which candidates commission polls?

|                               | (1)               | (2)               |
|-------------------------------|-------------------|-------------------|
| Final vote share              | 0.191<br>(0.031)  | 0.150<br>(0.048)  |
| Final vote share <sup>2</sup> | -0.105<br>(0.048) | -0.008<br>(0.066) |
| Log donations                 | 0.001<br>(0.001)  | 0.000<br>(0.001)  |
| Race fixed effects            | No                | Yes               |
| N (candidate-races)           | 8,515             | 8,515             |
| R <sup>2</sup>                | 0.018             | 0.429             |
| Mean of outcome               | 0.051             | 0.051             |

*Note:* Linear probability model of *ever self-sponsored* (1 if the candidate paid for at least one of their own polls; mean 0.052) on candidate-level covariates. Column (1) pools across races; column (2) adds municipality (race) fixed effects. Final vote share is in  $[0, 1]$ . Log donations is  $\log(1 + \text{total campaign donations})$  in BRL, summed from TSE 2024 *receita* filings. Standard errors clustered at the municipality.