Supporting Information Text

1. Continuity of Observations

Regression Discontinuity Designs (RDD) rely on the assumption that potential outcomes are continuously distributed at the treatment cutoff, with treatment assignment relying only on the running variable – in this case the winning margin for women candidates. That is, treatment assignment should be orthogonal to any confounding variable that may affect violence against women (VAW). The RDD assumption could be violated if treatment assignment is not orthogonal to any variable that may affect the outcome or if mayoral candidates can influence their assignment-to-treatment (the margin of victory) and sort nonrandomly around the threshold. We employ three sets of tests to provide evidence that the continuity assumption is met: formal tests of sorting, balance tests for the continuity of covariates around the threshold, and a placebo test with past VAW outcomes, specifically homicides of women and young women in 2010 and 2017. We select 2010 because it is the same year as the sociodemographic data we use in our analysis and 2017 because potential candidates for the 2018 election had to register their candidacy in late 2017. Null results found in each of the robustness checks indicate the continuous potential outcomes assumption is met and that treatment assignment is orthogonal to other variables that could affect VAW.

Sorting Tests.

McCrary Test. The RDD assumption would be violated if mayoral candidates can influence their assignment-to-treatment (the margin of victory) and sort nonrandomly around the threshold. In order to formally verify that there is no candidate sorting around the treatment cutoff, we conduct a standard McCrary test (1) and present the results here. This test uses the same RDD framework to explore outcomes around the cutoff but uses the density of observations as the outcome rather than the primary VAW outcomes of interest used for the main analysis. If the density of observations is discontinuous around the threshold, 0% margin of victory, then the assumption of continuous potential outcomes is violated. Figure S1 shows the binned number of observations below and above the 0% margin of victory threshold, where observations above the threshold indicate municipalities where a woman candidate beat a man, with the density of observations on either side of the threshold overlaid (and 95% confidence intervals). Although there is a small "jump" with a lower density of observations located above the threshold, this discontinuity is not statistically significant. There does not appear to be any identifiable sorting below or above the threshold. According to the formal test, the log difference in density height is -0.1744 (binwidth 0.01529) with a p-value of 0.31. This null result suggests that the continuity assumption is likely to hold in our research context.

Nonparametric Test. To provide further evidence, we also validate the continuity of observations using a nonparametric test from Cattaneo et al. (2020) (using the `R` package `rddensity`) that does not require binning (2). The nonparametric test (using jackknife standard errors) also indicates no evidence of sorting around the cutoff ($t = -0.52$; $p = 0.60$; effective $n = 415$).

Covariate Balance. The RDD assumption could be violated if the treatment assignment is not orthogonal to a variable that may affect VAW, and thus, confounding variables should be continuous around the cutoff. Using data from the 2010 Census on municipality-specific sociodemographic factors, including gender-specific variables such as number of women, women-run households, economically active women, and the average education of women, we conduct balance tests by estimating the RDD with these sociodemographic variables as outcomes. We estimate the RDD following the same procedure as the main RDD results (using the `R` package `rddensity`). Table S1 shows the results. We find no discontinuity at the threshold for any of the sociodemographic variables. The findings support the assumption that treatment assignment is orthogonal to other confounding characteristics of municipalities that may influence VAW.

Table S1. Covariate balance: Women politicians and demographic variables, RDD estimates.

Outcome	Estimate	SE	p	Bandwidth	Polynomial	Obs
Population	-9,455.16	31,279.10	0.76	0.12	1	309
Population, men	-4,778.35	15,275.12	0.75	0.12	1	307
Population, women	-4,780.73	16,028.79	0.77	0.12	1	311
Mean years of schooling	0.10	0.34	0.76	0.16	1	378
Mean years of schooling, men	0.02	0.34	0.96	0.17	1	378
Mean years of schooling, women	0.18	0.35	0.62	0.15	1	364
Economically active population	-4,426.67	13,645.66	0.75	0.12	1	307
Economically active population, men	-3,021.14	8,459.54	0.72	0.14	1	334
Economically active population, women	-1,766.38	5,095.26	0.73	0.12	1	304
Number of homes	-2,748.78	7,873.79	0.73	0.12	1	309
Number of homes, men head	-2,153.40	5,774.00	0.71	0.14	1	338
Number of homes, women head	-1,104.42	2,131.99	0.60	0.11	1	292
Married pop, 12 y/o and older	-4,439.33	12,841.54	0.73	0.12	1	309
Catholic population	-9,066.02	26,501.22	0.73	0.11	1	294

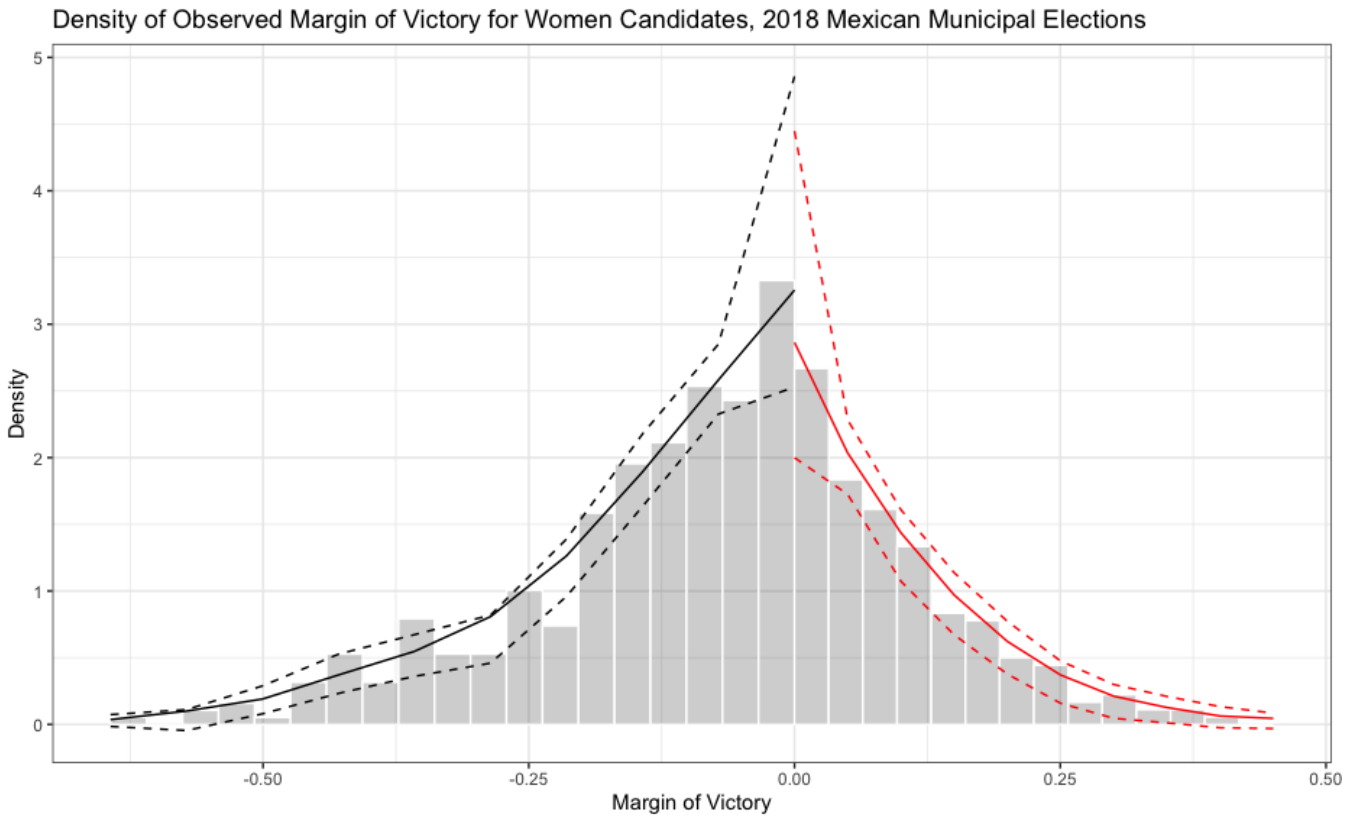

Fig. S1. Distribution of Margin of Victory Around the Threshold

Placebo Test with Past VAW Outcomes. We also use a placebo test to provide further evidence addressing two concerns: (1) that women politicians are self-selecting and winning close elections in municipalities with high VAW levels and (2) that a spurious correlation due to some third confounder is driving both VAW and the electoral success of women politicians in close elections.

We use homicides of women and young women in two separate years: 2010 (same year as the census that we use for the covariate balance tests, used for consistency) and 2017 (year before the elections when politicians are required to register as candidates). Data from both years would tell us whether women politicians are self-selecting and winning close elections in municipalities with high VAW levels. However, the placebo test using data from 2017 is most relevant because potential 2018 election candidates must register themselves at the end of 2017.

We estimate the RDD with these measures as outcomes and using the same specification as the main results (using the `R` package `rddensity`). Results are shown in Tables S2 and S3. We find that that electing a woman candidate in 2018 has no effect on these *past* VAW outcomes. This provides compelling additional evidence that women are not self-selecting into electoral races in municipalities that are particularly dangerous (or safe) for women or electorally beneficial for women, and lends support to the as-if-random assumption.

Table S2. Women politicians and homicides of women in 2010, placebo test, RDD estimates.

Outcome	Estimate	SE	p	Bandwidth	Polynomial	Obs
Homicides of women	0.05	0.10	0.61	0.07	1	206
Homicides of women	0.02	0.13	0.87	0.09	2	255
Homicides of young women	-0.003	0.06	0.96	0.07	1	207
Homicides of young women	-0.02	0.09	0.81	0.11	2	280

Table S3. Women politicians and homicides of women in 2017, placebo test, RDD estimates.

Outcome	Estimate	SE	p	Bandwidth	Polynomial	Obs
Homicides of women	-0.79	0.80	0.33	0.08	1	238
Homicides of women	-1.43	1.16	0.22	0.09	2	255
Homicides of young women	-0.44	0.53	0.41	0.08	1	225
Homicides of young women	-0.80	0.74	0.28	0.09	2	248

2. Spatial Distribution of Treated and Control Units

A related but different concern may be spatial sorting, i.e. that women politicians only win close elections in certain regions. In order to demonstrate that there is no spatial sorting of treated versus untreated municipalities, Figure S2 shows the geographic distribution of our sample. Of the 1,324 municipalities we collected data on, 559 (42%) held elections where a woman and a man were the top two vote-receiving candidates. Municipalities where elections took place in which a woman candidate defeated a man are shown in light blue and where a man candidate defeated a woman are shown in green. Municipalities where both candidates were the same gender are white and not included in the RDD estimations[*]. The map reveals that municipalities with a woman mayor are not spatially clustered in any particular region of Mexico.

3. Coding Candidate Gender

The Mexican government provides the gender of the winning candidate that becomes mayor but the gender of candidates that do not win is not systematically collected or reported.[†] This information is crucial for our study because the identification strategy relies on comparing municipalities where a woman candidate barely defeats a man candidate and municipalities where a man candidate barely defeats a woman candidate. We thus hand-code the gender of the first and second place candidates in each election using information from each state's electoral agency. Here we describe the data collection procedure for this information.

We recruited two undergraduate research assistants (RAs) and provided them with the political party and number of votes for each first and second place candidate for all municipal elections in 2018 from Magar (2018) (3). This information also included the gender of the winning candidate. To identify the gender of the second place candidate, the RAs were instructed to search for the electoral results of each municipality in each state's electoral agency. This is because municipal election data is stored by each state's electoral agency, not the federal electoral agency. For each state, the RAs searched for the list of candidates that included their political party and electoral results and matched the official state election results to Magar's

[*] Municipalities in the state of Oaxaca are also excluded, as is standard in the literature, because hundreds of municipalities follow indigenous self-governance that uses different electoral rules. We also exclude municipalities in the states of Tabasco and Yucatán due to lack of data on the gender of candidates. Any states that did not hold elections in 2018 are similarly not included in the estimations.

[†] Some states do report this information, though rarely in a systematic manner, while others do not.

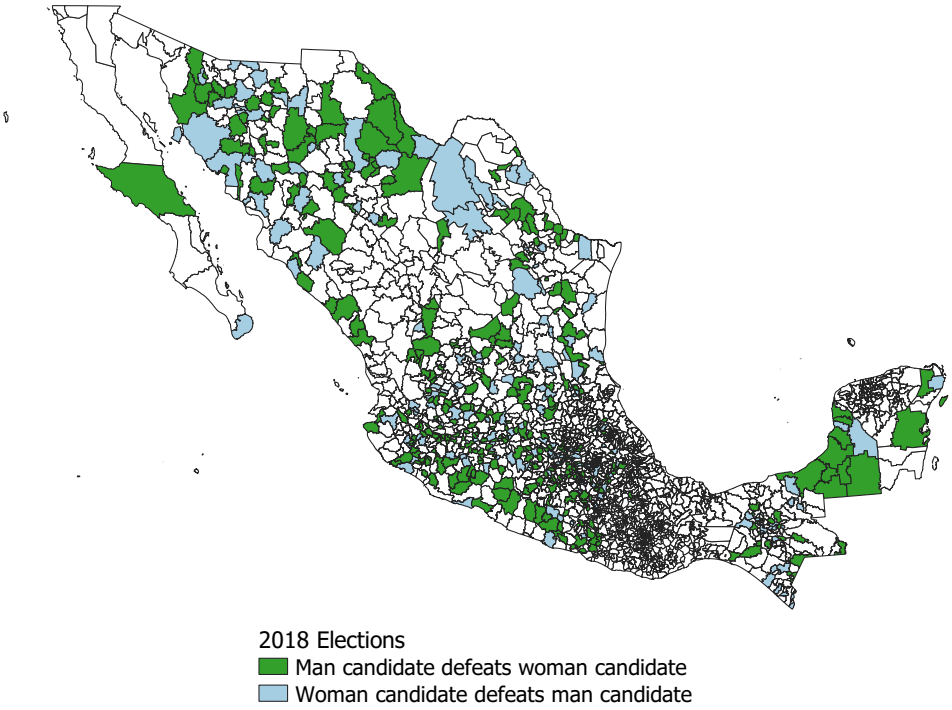

Fig. S2. Geographic distribution of municipalities with elections in 2018 where either a woman candidate defeated a man candidate (show in light blue) or a man candidate defeated a woman candidate (shown in green).

data. For each election, the RAs (1) identified the candidate that received the second most electoral votes and coded whether that candidate was a woman or a man based on their name, (2) verified that the first and second place candidates and their parties was accurate, and (3) verified that the number of votes for the first and second place candidates was accurate. The principal investigators (PIs), two of whom are of Mexican origin, trained the RAs and verified their work.

In Mexico, the vast majority of names are easily attributable to a gender. For names that are not gender specific, that the RAs could not code, or that the RAs were unsure about, they were instructed to leave blank spaces and highlight them for further review by the PIs. The PIs then went through the names the RAs could not identify and made coding decisions based on the name, and if the name was still unclear, determined their gender based on background research on each one of these unknown candidates. For example, the PIs routinely verified a candidate's gender through the candidate's personal campaign website or news stories covering the candidates.

4. RDD Results: Homicides of Men and Non-VAW Crimes

To investigate whether the effect of women politicians on VAW is a general effect on violence and crime or specifically about VAW, we also estimate the main RDD specification using non-VAW outcomes. Specifically, we estimate the RDD using outcome measures of other types of violence and crime: the homicides of men, homicides of young men, and four of the most prevalent crimes in Mexico (extortion, home burglary and vehicle theft, kidnapping, and drug dealing). Data on homicides comes from death certificate data from 2019 and 2020 and is collected from Mexico's National Institute of Statistics and Geography (*Instituto Nacional de Estadística, Geografía y Informatica* or INEGI) (4). Data on crimes is from the Executive Secretariat of the National Public Security System (*Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública* or SESNSP) (5), and measures reported crimes from 2019 - 2021. Results are shown in Table S5 and the descriptive statistics of these outcomes are shown in Table S4.

Table S4. Summary statistics: Homicides of men (2019-20) and non-VAW crimes (2019-21)

Statistic	Mean	St. Dev.	N
Homicides of men	17.43	53.83	559
Homicides of young men	13.06	41.95	559
Extortion	4.89	17.70	559
Theft	124.18	477.16	559
Drug dealing	29.16	122.00	559
Kidnapping	0.46	1.33	559

First, we find that women politicians that win narrow elections have a short-term (first year in office) negative effect on homicides of men and young men, but that these effects become smaller and lose their statistical significance at the 5% level during a woman politician's second year in office. The effect sizes for the first year of a woman's administration are similar in size across homicides of women and men. That is, the point estimates are all close to the mean number of homicides of their respective measures. This tells us that the effects are substantively large the first year for homicides of both men and women, but become larger during the second year for homicides of women and remain statistically significant, while the effects on homicides for men get smaller and lose their statistical significance at the 5% level during a woman politician's second year in office. This suggests that women have an overall effect on reducing homicides, but this effect is short-lived for homicides of men and not only persistent for homicides of women but even more pronounced in subsequent years.

Second, we also find that women politicians have no effect on the prevalence of reported non-VAW crimes for any year (no results are statistically significant at either $p < 0.05$ or $p < 0.1$). Interestingly, though not statistically significant, some point estimates for kidnapping, theft, and extortion are positive, suggesting that the consistent negative effects when using VAW crimes is not due to some phenomenon wherein all crimes and forms of violence are lower in municipalities with women mayors. This provides strong evidence that women politicians are having an effect on VAW outcomes specifically.

Together, these results suggest that women politicians reduce VAW crimes - particularly severe forms of VAW - and not crimes in general, though they do have some short-term effect on homicides of men as well.

5. Multi-Cutoff RDD

Since our RDD design is based on winning margins from a plurality electoral system, we are estimating a multi-cutoff RDD (6). In other words, unlike a single-cutoff RDD where the cutoff is the same for all units, the cutoff in plurality elections depend on the vote share of each candidate. For example, one candidate could win with 34% of the vote to an opponent with 30% of the vote, while another candidate could win with 59% of the vote to an opponent with 40% of the vote. By using the margin of victory as our running variable, we are normalizing the running variable and pooling our units. By doing so, our RDD is estimating the weighted average of the local average treatment effect across vote shares (6). Our coefficient of interest is this pooled estimand. We thus focus on estimating the pooled estimand and leave heterogeneity unexplored. We also do not explore heterogeneity due to our relatively small sample size.

However, if we assume constant treatment effects, our RDD estimate can be interpreted like a single-cutoff RDD design: the overall average of the average treatment effects (6). To assess this assumption, Cattaneo et al. (2016) recommended plotting

Table S5. Women politicians and Non-Vaw Outcomes, RDD estimates.

Outcome	Year	Estimate	SE	p	Bandwidth	Polynomial	Obs
Homicides, men	2019	-18.54	7.74	0.02	0.06	1	176
Homicides, men	2020	-13.11	6.87	0.06	0.07	1	206
Homicides, young men	2019	-13.91	6.02	0.02	0.06	1	182
Homicides, young men	2020	-10.60	5.50	0.05	0.07	1	199
Extortion	2019	0.22	3.37	0.95	0.10	1	274
Extortion	2020	-2.05	4.51	0.65	0.18	1	404
Extortion	2021	-2.06	5.33	0.70	0.14	1	339
Theft	2019	-0.33	67.68	0.80	0.10	1	264
Theft	2020	17.11	67.68	0.80	0.10	1	264
Theft	2021	15.11	65.36	0.82	0.10	1	266
Drug dealing	2019	-6.29	18.20	0.73	0.08	1	225
Drug dealing	2020	-24.30	24.85	0.33	0.08	1	220
Drug dealing	2021	-40.10	26.96	0.14	0.07	1	210
Kidnapping	2019	0.02	0.37	0.96	0.10	1	264
Kidnapping	2020	0.13	0.23	0.57	0.09	1	248
Kidnapping	2021	0.04	0.24	0.86	0.08	1	218

the vote share of the second place candidate ("the strongest opponent"). As noted by these scholars, "if most of the mass in the distribution is near the same cutoff value, then the analyst can treat the design as equivalent to a single-cutoff RD design" (6, p. 1246). Figures S3 and S4 plot these distributions for the full sample and for close elections (winning margin $< 10\%$). The figures show a unimodal distribution centering around 30%. This makes intuitive sense in the Mexican case, as Mexico had three major parties in 2018: Morena, PAN, and PRI. We take this as suggestive evidence that we may be able to treat our design as equivalent to a single-cutoff RD design and interpret the RDD coefficients as the local average treatment effects.

6. Deviations from Pre-Analysis Plan

In this section we explicitly identify and explain deviations we took from the pre-analysis plan (PAP) we registered prior to data collection and that is available at the Open Science Foundation Registry: https://osf.io/7ty4q.

First, we only specified that we would use covariate balance tests to check the RDD continuity assumption. In addition to the covariate balance tests we registered, we chose to also use formal sorting tests. We chose to do this because it has become standard practice and to provide even more robustness than the original PAP anticipated. An additional robustness check we run is a placebo test using past outcomes as the dependent variable. We did not register this test in the PAP. Again, this is an additional robustness check that provides further credibility to the main results.

Second, in our pre-registered research design, we noted that we planned to estimate the RDD using two procedures to calculate optimal bandwidths: Imbens and Kalyanaraman (2012) (herein IK) (7) and Calonico, Cattaneo, and Farrell (2020) (herein CCF) (8). However, CCF improves upon the MSE-optimal bandwidth selectors from IK, as discussed in (9), and we therefore only calculate bandwidths using this method.

Third, in the PAP we specified that data on local elections would come from a third-party repository. However, after beginning the data collection on the gender of candidates we noticed that some of the election results were not completely accurate. This is likely because election results take time to verify. We therefore collected and verified each election result directly from each state's electoral agency.

Fourth, we only planned to run the RDD using VAW measures as outcomes. When the question arose as to whether women politicians affect other forms of violence and crimes that are not gendered, we decided to run additional analyses using homicides of men and non-VAW crimes (extortion, theft, kidnapping, and drug dealing). These tests were not included in the PAP. Nevertheless, we decided to estimate the RDD using these outcomes following the research design we *had* registered so as to not deviate from the original plan. It should also be noted that these are additional tests and not main results.

Finally, the PAP included a preliminary plan to explore heterogeneous effects. However, after collecting election and gender data we decided that the sample size was likely not large enough to give us the power to conduct these tests. We therefore decided against collecting additional data.

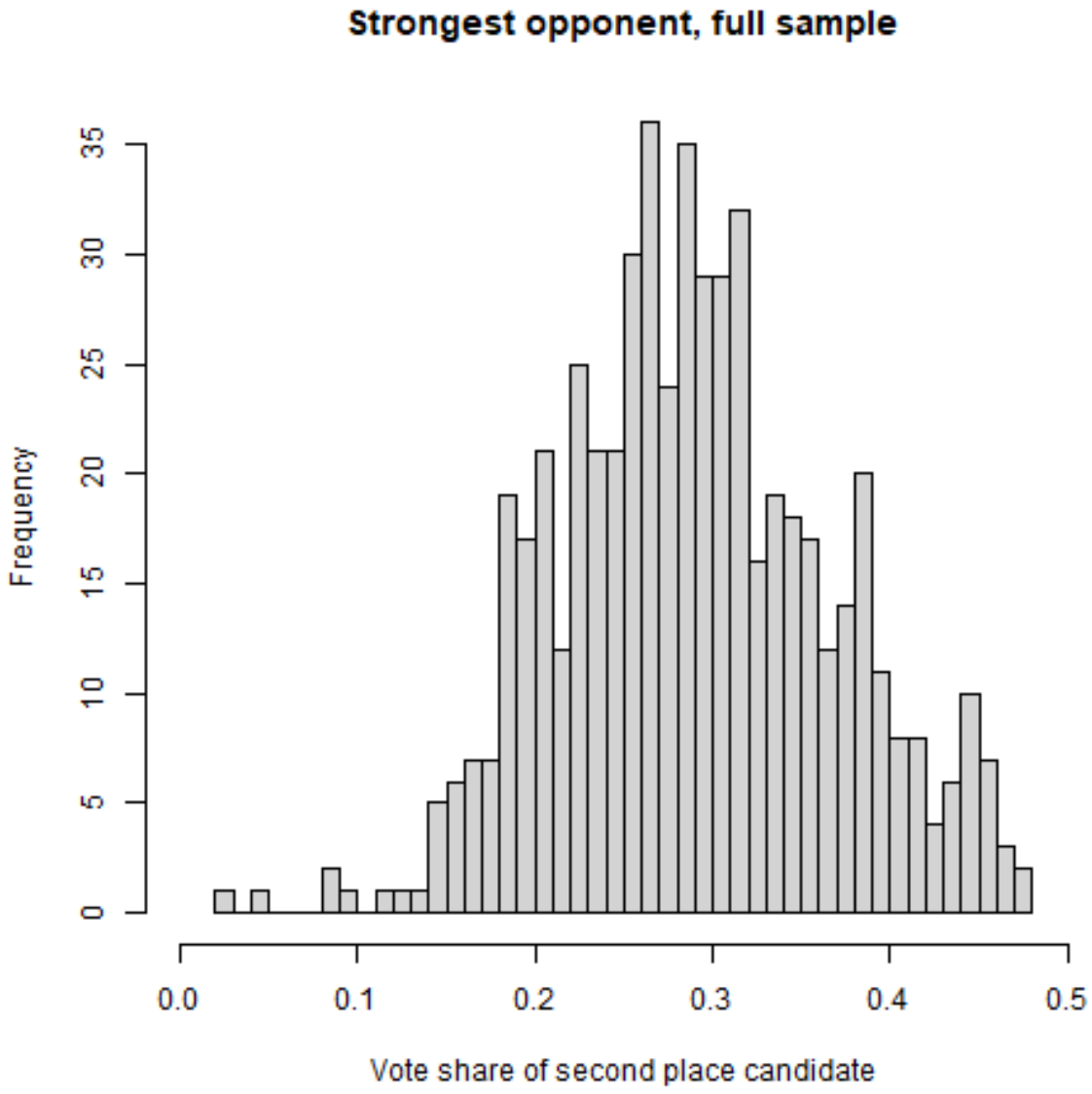

Fig. S3. Histogram of vote share of second place candidate in the 2018 local elections in Mexico.

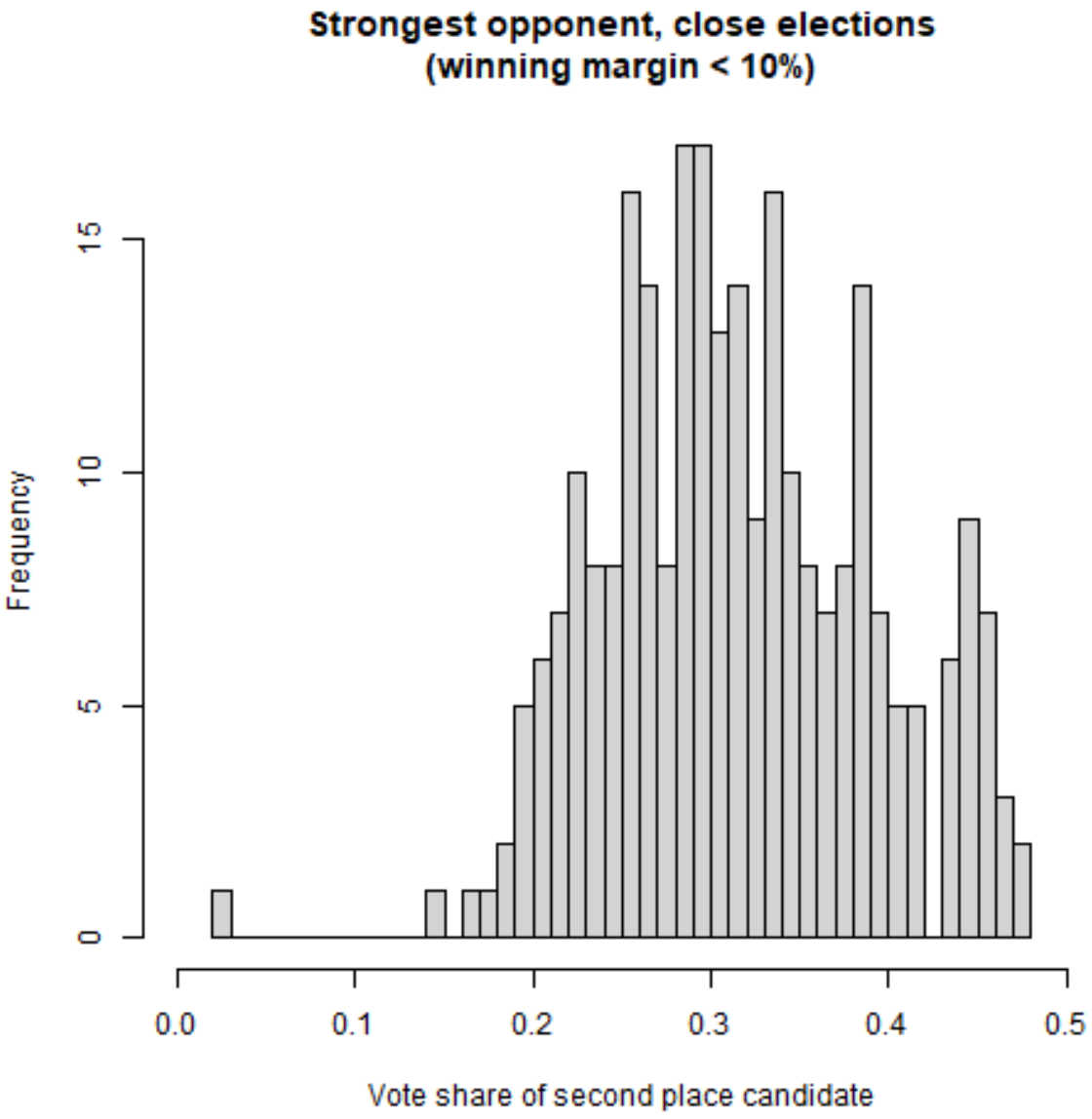

Fig. S4. Histogram of vote share of second place candidate in close elections during the 2018 local elections in Mexico.

SI Dataset S1 (`vaw_data.csv`)
Dataset used for the main results and robustness checks.

References

1. J McCrary, Manipulation of the running variable in the regression discontinuity design: A density test. *J. Econom.* **142**, 698–714 (2008).
2. MD Cattaneo, M Jansson, X Ma, Simple local polynomial density estimators. *J. Am. Stat. Assoc.* **115**, 1449–1455 (2020).
3. E Magar, Recent mexican election vote returns repository (2018).
4. Instituto Nacional de Estadística, Geografía y Informatica, Mortalidad. (2021).
5. Secretariado Ejecutivo del Sistema Nacional de Seguridad Pública, Incidencia delictiva. (2022).
6. MD Cattaneo, R Titiunik, G Vazquez-Bare, L Keele, Interpreting regression discontinuity designs with multiple cutoffs. *The J. Polit.* **78**, 1229–1248 (2016).
7. G Imbens, K Kalyanaraman, Optimal bandwidth choice for the regression discontinuity estimator. *Rev. Econ. Stud.* **79**, 933–959 (2012).
8. S Calonico, MD Cattaneo, MH Farrell, Optimal bandwidth choice for robust bias-corrected inference in regression discontinuity designs. *The Econom. J.* **23**, 192–210 (2019).
9. S Calonico, MD Cattaneo, R Titiunik, Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* **82**, 2295–2326 (2014).