

Supplementary Information

From Olympic Gold to Government Approval?:

Evidence of Limited Misattribution from the 2024 Paris Games

January 27, 2026

Contents

A Summary of Existing Research	3
B Survey Questionnaire	8
B.1 Outcome variables	8
B.2 Covariates	8
C Full Equivalence-Test Results: Feeling Thermometer	10
D Robustness Checks with Unexpected Events during Survey Design Approach	14
D.1 Demographic issues	15
D.2 Balance checks and results to address imbalances	21
D.3 Visualizing Rolling RDD Results	28
D.4 No response analysis	33
D.5 Time trends	34

A Summary of Existing Research

Table A1: Summary of existing literature

Results	Author	Country	Method	External shocks	Outcome	Effects	
Positive & Negative significant effects on incumbent support	Achen and Bartels (2004), Achen et al. (2018)	United States	Multivariate regression	Shark attacks	Incumbent share	vote	The shark attack reduced support for the incumbent in coastal New Jersey
	Healy et al. (2010)	United States	Multivariate regression	American football/Basketball	Incumbent share	vote	Local team wins increase support for incumbents, while losses decrease it
	Chang and Berdiev (2015)	156 countries	Conditional logit	General disasters	Government placement	re-	Natural disasters cause government replacement
	Gallego (2018)	Colombia	DID	Rainfall	Reelection		Incumbents' reelection chances increase through clientelism after disaster
	Masiero and Santarossa (2021)	Italy	DID	Earthquake	Incumbent election and share	re-vote	Earthquake occurrence enhances the reelection probability and vote share
	Cooperman (2022)	Brazil	Regression	Drought (declaration)	Reelection		Incumbents who declare drought are more likely to be re-elected regardless of actual rainfall
	Visconti (2022)	Chile	Natural and conjoint experiment	Flood	Ideological voting		Respondents with material damage are more likely to support a left-wing party or an independent

Continued on next page

Table A1 – *Continued from previous page*

Results	Author	Country	Method	External shocks	Outcome	Effects
	Graham et al. (2023)	United States	Replication of Achen and Bartels (2016) & Healy et al. (2010)	Floods, droughts, tornadoes & football games	Incumbent support	Re-examines previous findings on disasters and football; finds limited and fragile effects on incumbent support
	Fukumoto and Kikuta (2024)	Japan	IV	Rainfall	Incumbent share	Disaster relief increases the incumbent vote share
Mixed effects	Gaspar and Reeves (2011)	United States	Panel regression	Weather damage	Incumbent share	Voters decrease support for incumbents due to natural disasters, but attentive voters support incumbents who focus on reconstruction
	Cole et al. (2012)	India	Regression	Rainfall	Incumbent share	Rainfall decreases the incumbent vote share, but incumbents who rigorously address disasters increase their vote share
	Hilbig and Riaz (2024)	Germany	RDD	Flood	Green party support	Voters demand disaster relief rather than climate change prevention (public goods), thus support green parties

Continued on next page

Table A1 – *Continued from previous page*

Results	Author	Country	Method	External shocks	Outcome	Effects
	Kondoh and Miyazaki (2024)	Japan	Multivariate regression	General disasters	Incumbent share	vote Natural disasters do not increase support in the short term, but over the long term they increase support not through relief, but through political alignment
Null effects	Fowler and Pablo Montagnes (2015)	United States	DID	College football	Two-party vote	College football effects are false positives (if college football games truly influenced elections, effects should be strongest where interest is highest, but they find the opposite pattern)
	Fowler and Hall (2018)	United States	DID	Shark attacks	Two-party Democratic vote share	Shark attacks have a null effect on incumbent support
	Rapeli and Söderlund (2022)	Finland	Panel regression	Olympic events	Government parties' support/Prime ministerial support	Olympic medal events do not affect incumbent support in a multiparty system
	Müller and Kneafsey (2023)	Ireland	Multivariate regression & robustness checks with the UESD	Win/loss of inter-county senior men's Gaelic football and hurling matches	Support for candidates as the percentage-point difference in first-preference vote shares between election t and election $t - 1$	Even using UESD robustness checks, sporting games do not influence incumbent support

Continued on next page

Table A1 – Continued from previous page

Results	Author	Country	Method	External shocks	Outcome	Effects
	Catalinac and Mu-raoka (2023)	Japan	GRDD	Snow	Incumbent share	Programmatic snow subsidies do not enhance incumbent vote share

Note: DID is difference-in-differences method. IV is instrumental-variable method. RDD is regression discontinuity design. UESD is unexpected-events-during-survey design. GRDD is geographical regression discontinuity design.

B Survey Questionnaire

B.1 Outcome variables

- **Government approval:** *Do you approve the Fumio Kishida administration?*

Approve / Disapprove / Neither approve or disapprove

- **Feeling thermometer:** *I'd like to get your feelings toward some of our party leaders and political group who are in the news these days. Ratings between 50 degrees and 100 degrees mean that you feel favorable and warm toward the person. Ratings between 0 degrees and 50 degrees mean that you don't feel favorable toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person.*

Liberal Democratic Party / Constitutional Democratic Party of Japan / Japan Innovation Party / Japanese Communist Party / Kishida, Fumio / Izumi, Kenta / Baba, Nobuyuki / Tamura, Tomoko / Japanese government

- **Economic evaluation:** *Please select the option that best represents your opinion on the current Japanese economic condition and your own living conditions.*

Very good / Good / Neither good nor bad / Bad / Very bad

B.2 Covariates

- **Party support:** *Which party do you usually support?*

Liberal Democratic Party / Constitutional Democratic Party of Japan / Japan Innovation Party / Komeito / Japanese Communist Party / Democratic Party for the People / Reiwa Shinsengumi / Social Democratic Party / Party for Realization of Education without Degeneration / Party to Protect the People from NHK / Sanseito / Other (please specify) / No such party / Don't know / Refuse to answer

- **Olympic interest:** *How interested are you in the Paris Olympics, which will begin on July 27, 2024?*

Very interested / Somewhat interested / Not very interested / Not interested at all

- **Watch the ceremony:** *Did you watch the opening ceremony of the Paris Olympics?*

Watched in progress on TV, etc. / Watched a recording, etc. / Did not watch

- **Patriotism and Nationalism:** *How do you think about the following opinions? Please choose one answer for each.*

Agree / Somewhat agree / Neither agree nor disagree / Somewhat disagree / Disagree

1. I take pride in being Japanese.
2. If I were to be reborn, I would choose to be Japanese again.
3. I have a deep love for Japan.
4. Given Japan's economic power, Japan should have a more significant position in international assemblies.
5. Japan's post-war economic growth is due to the diligent efforts of its people.
6. Japan's trade profits are a result of its advanced technology and hard work.

C Full Equivalence-Test Results: Feeling Thermometer

As in the main text, we visualise the equivalence-test results using heatmaps that place bin width on the horizontal axis and the equivalence-margin multiplier k on the vertical axis. For each cell, we report two bounds: the time-bin margin EB_{TB} , defined by $\varepsilon_{TB} = k \times SD(\text{bin means})$, and the cross-sectional margin EB_{CS} , defined by $\varepsilon_{CS} = k \times SD_{\text{pooled}}$. The panel header lists the 90% confidence interval for the treatment effect, defined as treated minus control, while the shading indicates whether that interval lies within both sets of bounds, only within the cross-sectional bounds, or within neither. Because the main text focuses on cabinet approval, the detailed equivalence-test results for the feeling thermometer are discussed in this section of SI C; as shown below, their implications closely mirror those for cabinet approval.

Feeling thermometer: overlap of equivalence (1st gold medal)

Notation: $EB_{TB} = \pm e_{TB}$ ($e_{TB} = k \times SD$ of bin means); $EB_{CS} = \pm e_{CS}$ ($e_{CS} = k \times \text{pooled } SD$). 90% CI (treated – control) = $[-0.618, 6.453]$

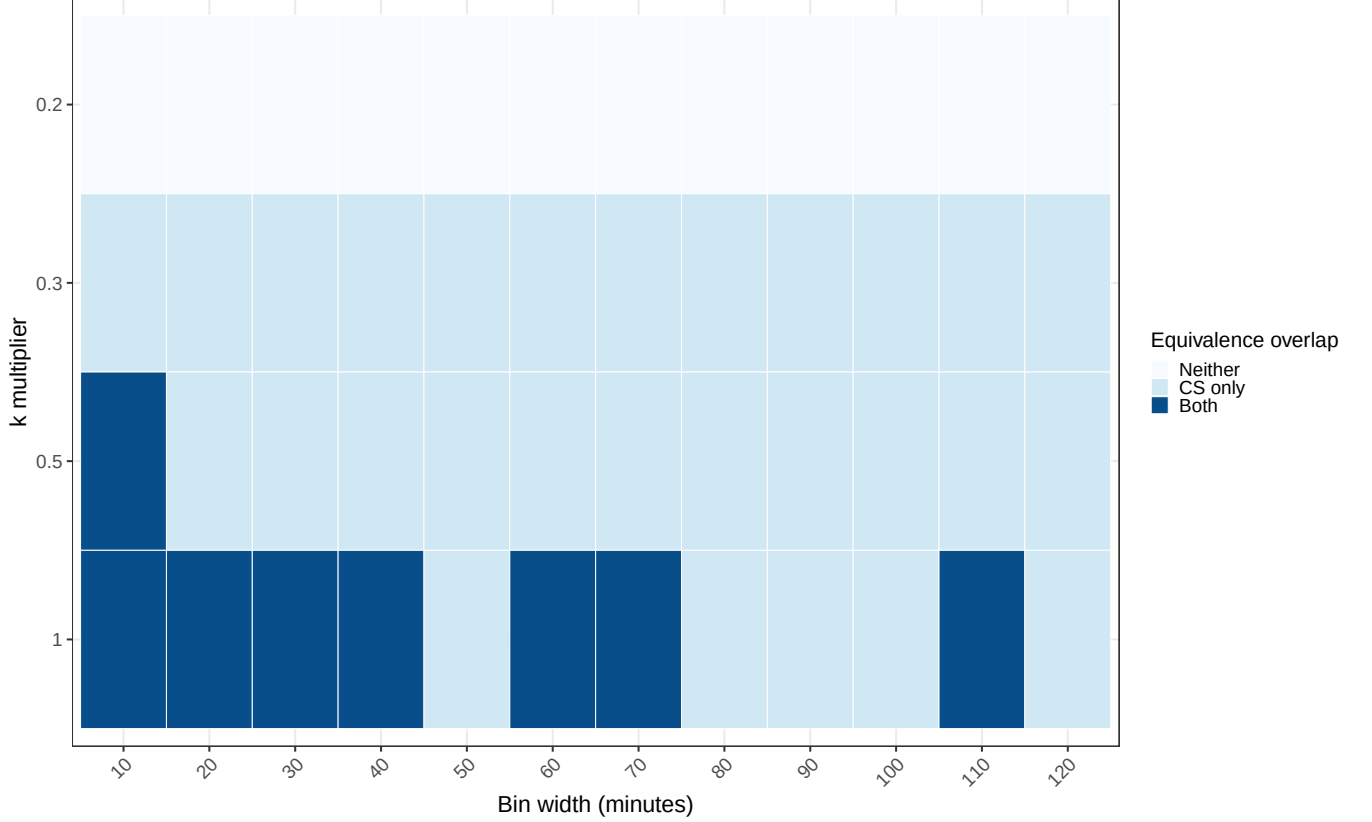


Figure C1: For the first gold medal, cross-sectional equivalence is missed at $k = 0.2$ but holds from $k = 0.3$ upward, and time-bin equivalence appears mainly at $k = 1.0$ with narrow bins, implying at most a small effect on the feeling thermometer

Note: This figure displays a heatmap of equivalence-test results for the feeling thermometer in response to the first gold medal during the observation window. Columns indicate the width of the time bins (30, 60, 90, and 120 minutes), and rows indicate the equivalence-margin multiplier k (0.2, 0.3, 0.5, and 1.0). For each cell, two equivalence bounds are reported: the time-bin margin EB_{TB} , defined by $\varepsilon_{TB} = k \times SD(\text{bin means})$, and the cross-sectional margin EB_{CS} , defined by $\varepsilon_{CS} = k \times SD_{\text{pooled}}$. The panel header reports the 90% confidence interval for the treatment effect (treated minus control), which for this outcome is $[-0.618, 6.453]$. Cell shading indicates whether that interval lies within both sets of bounds, only within the cross-sectional bounds, or within neither. See also *Note* in Figure 4 for the notation in the main text.

Feeling thermometer: overlap of equivalence (2nd gold medal)

Notation: $EB_{TB} = \pm e_{TB}$ ($e_{TB} = k \times \text{SD of bin means}$); $EB_{CS} = \pm e_{CS}$ ($e_{CS} = k \times \text{pooled SD}$).
 90% CI (treated – control) = $[-3.365, 3.256]$

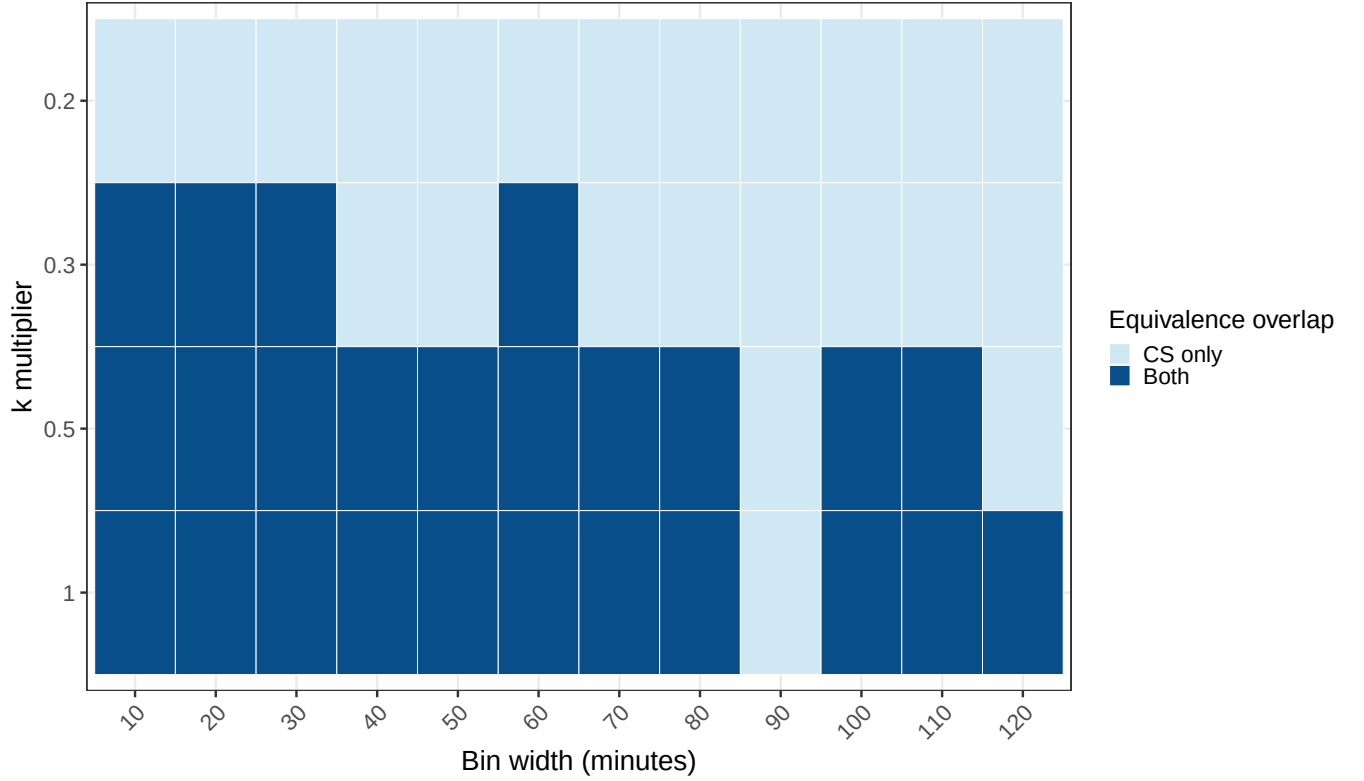


Figure C2: For the second gold medal, cross-sectional equivalence is missed at $k = 0.2$ but holds from $k = 0.3$ upward, and time-bin equivalence appears mainly at $k = 1.0$ with narrow bins, implying at most a small effect on the feeling thermometer

Note: This figure presents the results of equivalence tests for the feeling thermometer in response to the second gold medal during the observation window. The panel header reports the 90% confidence interval for the treatment effect (treated minus control), which for this outcome is $[-3.365, 3.256]$. See also *Note* in Figure C1.

For the feeling thermometer after the first gold medal, the 90% confidence interval $[-0.618, 6.453]$ narrowly exceeds the cross-sectional bound at $k = 0.2$, where the margin is about ± 5.224 points, yet it is comfortably contained from $k = 0.3$ upward, where the margin is about ± 7.836 points. Under the time-bin criterion, the interval typically exceeds EB_{TB} at $k = 0.5$, and both sets of bounds are met mainly at $k = 1.0$ with narrow bins. Consequently, large negative effects are ruled out, cross-sectional equivalence is supported from ± 0.3 SD, and time-bin equivalence remains out of reach unless tolerance is increased and bins are narrow, so the pattern is best characterized as an imprecise-null result relative to stricter time-bin margins.

For the feeling thermometer after the second gold medal, the 90% confidence interval $[-3.365, 3.256]$ is well contained within the cross-sectional ± 0.2 SD margin of approximately ± 5.055 points for every bin width, which supports a precise-null claim at that benchmark. Under the time-bin benchmark, both sets of bounds are already met at $k = 0.3$ when bins are narrow, and the region satisfying both expands as k increases. Therefore, even under conservative settings, the evidence indicates statistical equivalence and an effect that is essentially zero to very small.

D Robustness Checks with Unexpected Events during Survey Design Approach

D.1 Demographic issues

Figures D1—D2 compare the national population shares with the composition of the RDD samples for the first and second gold medals. Points lying on the 45-degree line would indicate perfect correspondence; in both panels, the markers cluster reasonably close to this line.¹ The RDD samples are somewhat older, slightly more male, and contain a higher share of university graduates than the national benchmarks, but the gaps are generally on the order of 10–15 percentage points rather than extreme divergences. In this sense, the late-night online respondents are not representative of the entire Japanese population, but their demographic profile does not appear so atypical as to make the analyses uninformative about broader patterns. However, further investigation is needed regarding the time-varying changes in demographics and the balance between the treatment and control groups.

¹Because we do not have access to micro data from other surveys conducted by the same firm, YCS, a strict side-by-side comparison would be desirable in principle but is not feasible, and we note this as a limitation.

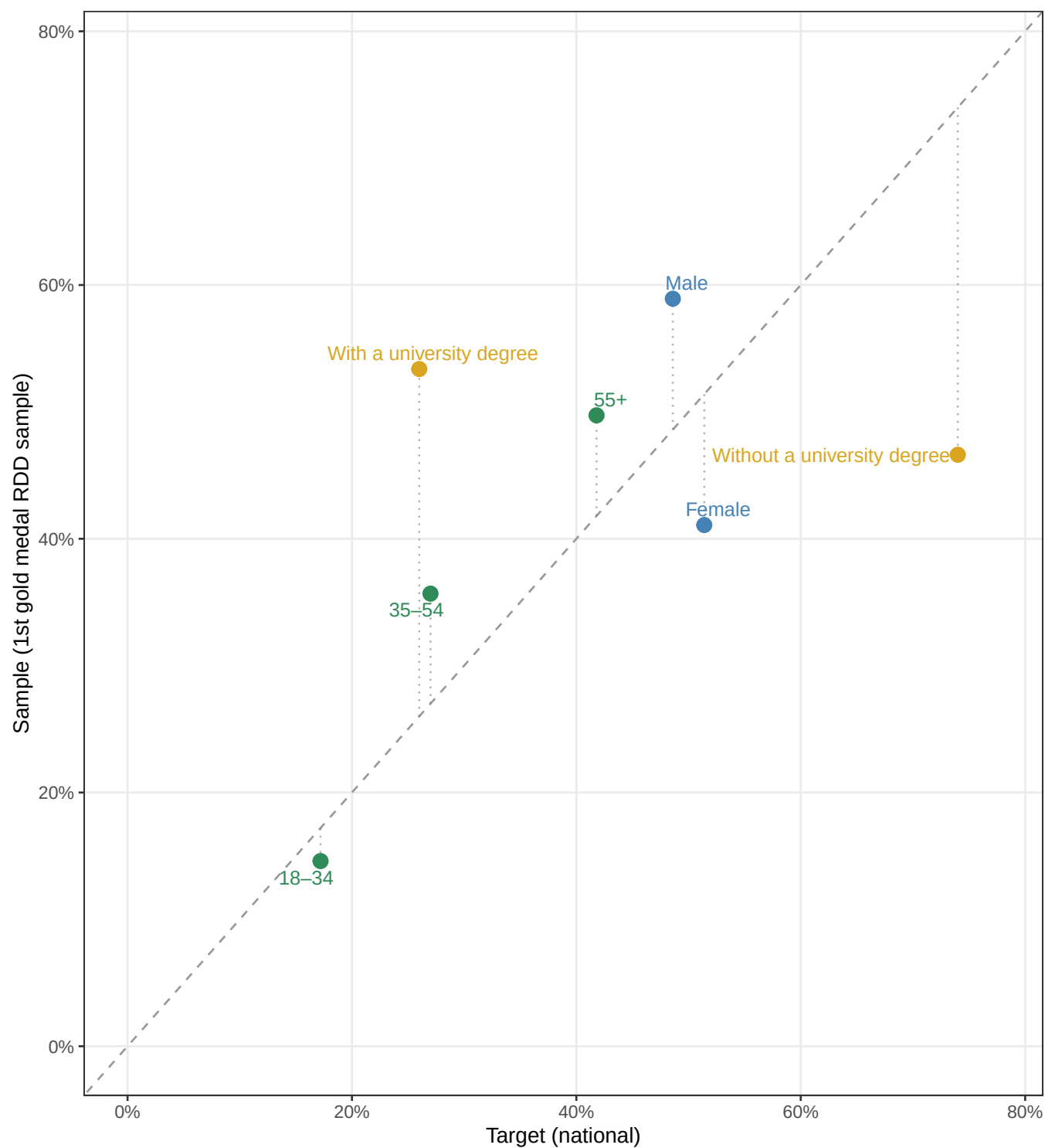


Figure D1: Comparison with national demography (1st gold medal)

Note: The figure compares national population (“target”) shares with observed shares in the RDD sample for the first gold-medal analysis. The horizontal axis shows population target shares and the vertical axis shows observed sample shares (percent). Each point corresponds to a demographic category (gender: female/male; age: 18–34, 35–54, 55+; education: with vs. without a university degree). The 45° diagonal is a reference line; points closer to the line indicate better alignment between the sample and the population. No confidence intervals are shown because the figure summarizes simple proportions.

FigureD3—D4 illustrate how the sample composition changes over the course of the survey night. As the fieldwork moves from shortly after midnight toward the early morning, respondents become increasingly female, older, and somewhat lower-income, while the share of university graduates declines. In other words, the demographic composition of our late-night online sample is not temporally stable, and it clearly drifts away from the attributes observed at the beginning of the survey period.

To quantify this pattern more precisely, we estimate linear regressions that take each demographic characteristic as the dependent variable and include as regressors a dummy for having responded after the medal cutoff and the running variable (survey time). For the first gold medal in the women’s 48kg event, the share of women increases by about 0.14 points at the cutoff (coefficient on cutoff = 0.14, with a 95% confidence interval of approximately [0.04, 0.24]), and the average age rises by roughly 9 years (9.1, 95% CI [5.9, 12.3]). By contrast, the income category decreases by about 1.2 steps (−1.20, 95% CI [−1.9, −0.5]). The point estimate for education is modestly negative (−0.18), but its 95% confidence interval includes zero. These results confirm that, at least around the first medal, the sample shifts toward older, more female, and somewhat lower-income respondents.²

By contrast, for the second gold medal in the men’s 66kg event we do not observe such discontinuities. For gender, age, education, and income alike, the coefficients on cutoff are very small in magnitude and statistically indistinguishable from zero (for example, the coefficient for the female share is 0.01 with a 95% confidence interval of approximately [−0.06, 0.09], and the coefficient for age is −1.19 with a 95% confidence interval of [−3.6, 1.2]). For this second event, the demographic composition follows a smooth time trend, but there is no discrete jump at the moment when the medal result becomes known.

These analyses corroborate that our late-night online respondents constitute a selective convenience sample and that, especially around the first medal, sample composition is not perfectly balanced before and after the cutoff. At the same time, the absence of sizable shifts in composition around the second medal suggests that the null result for this event is unlikely to be driven solely by changes in who responds to the survey. Moreover, all of our main RDD specifications condition on gender, age, education, and income, so the estimated treatment effects are identified from comparisons among respondents with similar observed characteristics who happened to answer just before versus just after each medal. We therefore interpret our findings as indicating that any immediate effect of these judo gold medals on cabinet approval was, at most, limited within the late-night online population observed in this study, while acknowledging that we cannot generalize our estimates to groups that are under-represented in our data—such as blue-collar workers who are unlikely to be awake and responding to online surveys at these hours.

²For full OLS results, see the corresponding section of the replication codes.

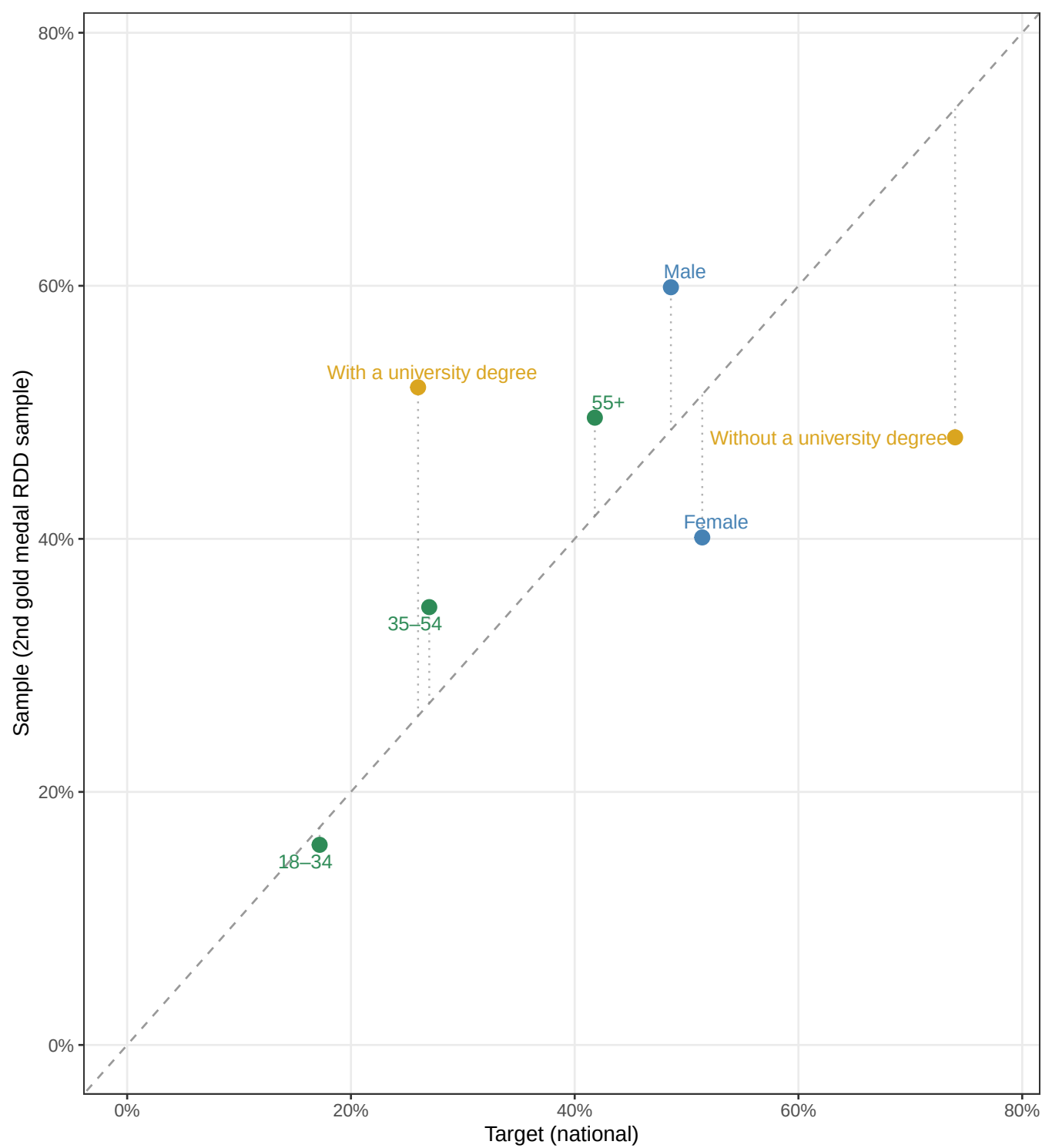


Figure D2: Comparison with national demography (2nd gold medal)

Note: The figure compares national population ("target") shares with observed shares in the RDD sample for the first gold-medal analysis. See also *Note* in Figure D1.

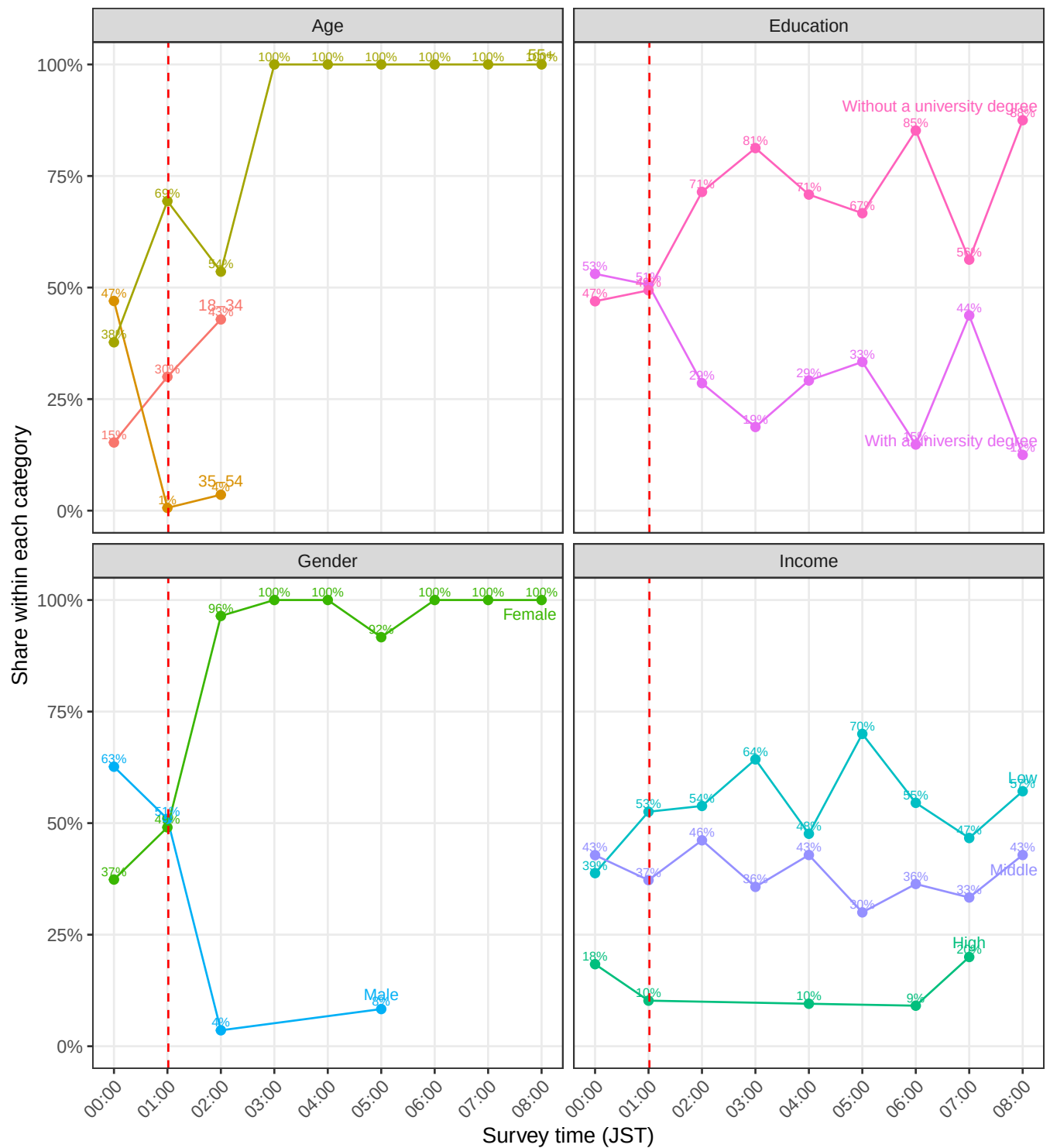


Figure D3: Sample characteristics tend to diverge substantially between the pre- and post-treatment periods (1st gold medal)

Note: The figure plots how the demographic composition of the RDD sample evolves over the survey night for the first gold-medal analysis. Each panel corresponds to one attribute (Age, Education, Gender, Income) and shows, by one-hour survey-time bins (JST), the share of respondents within each category (for example, ages 18–34, 35–54, 55+; with vs. without a university degree; female vs. male; low/middle/high income). Percentages next to the lines indicate the proportion (in percent) of each subgroup at a given time. The vertical dashed line marks the time at which the first gold medal was confirmed. No confidence intervals are displayed, as the figure summarizes descriptive proportions.

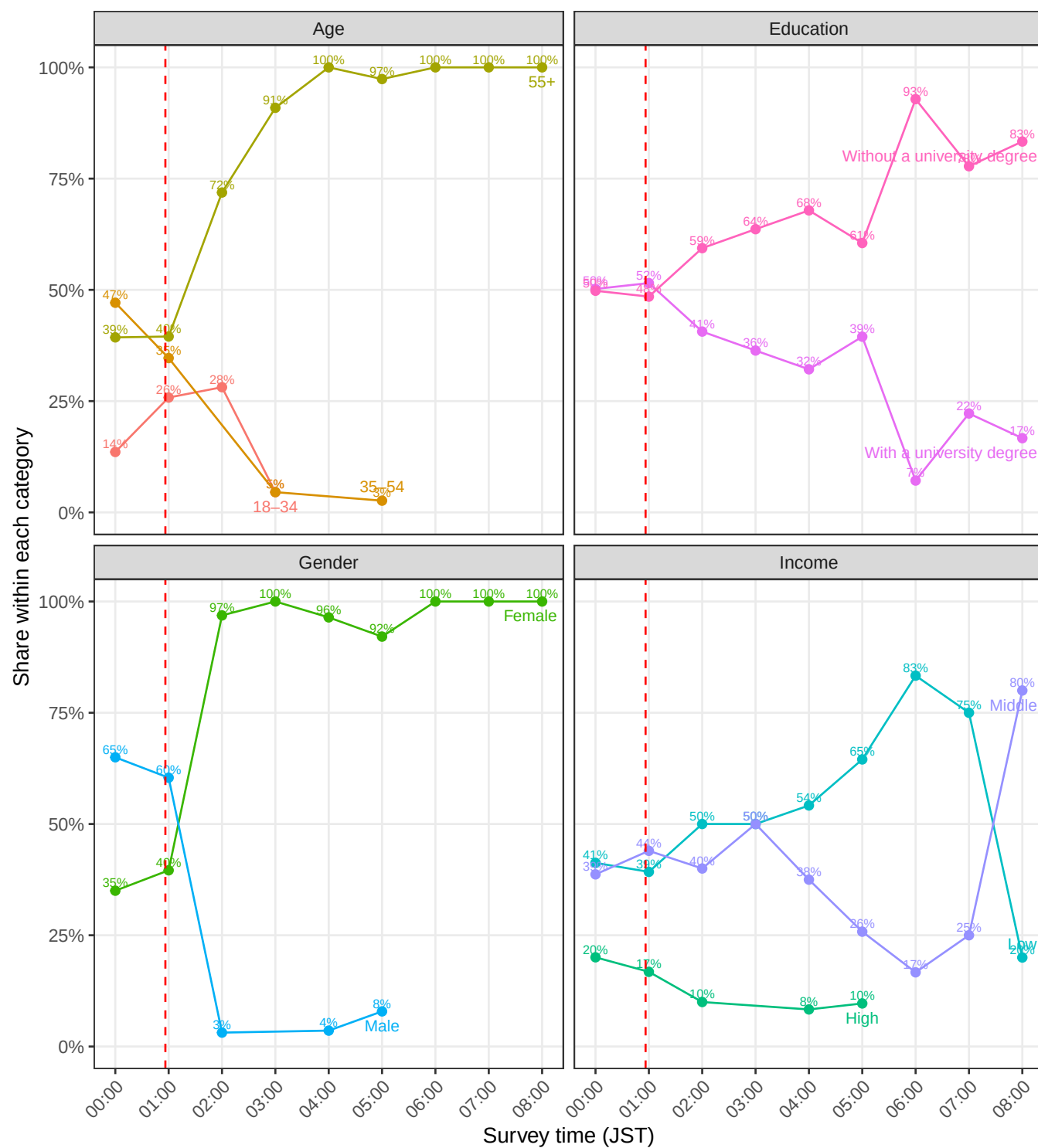


Figure D4: Sample characteristics tend to diverge substantially between the pre- and post-treatment periods (2nd gold medal)

Note: The figure plots how the demographic composition of the RDD sample evolves over the survey night for the second gold-medal analysis. See also *Note* in Figure D3.

D.2 Balance checks and results to address imbalances

Balance check

We conducted a balance check by comparing the means and standard deviations of various covariates between the control and treatment groups. These covariates include age, gender (female), education, income, and ruling parties support. The underlying assumption is that, in the absence of any treatment effect, these covariates should be similarly distributed across both groups. Any significant differences in these covariates could indicate an imbalance, which might confound the results.

Treatment	Control (Before)			Treatment (After)			
Variable	N	Mean	SD	N	Mean	SD	Test
Age	558	50	14	182	63	17	$F = 118.785^{***}$
Female	558	0.32	0.47	182	0.69	0.46	$F = 88.224^{***}$
Education	558	3.3	0.91	182	3	0.92	$F = 15.873^{***}$
Income	558	7.2	3.3	182	6	2.8	$F = 18.208^{***}$
In-partisans	558	0.3	0.46	182	0.32	0.47	$F = 0.2$

Table D1: Balance check results represent a clear imbalance between the treatment and control groups (The first gold medal case)

Note: This table presents the results of a balance check conducted using summary statistics, generated by the `sumtable` function in `vtable` package. The table compares key demographic and political support variables between the control group (Before) and the treatment group (After) to ensure that these groups are comparable before and after the treatment, which in this context refers to the occurrence of an event or intervention. The table includes five variables: Age, Female (gender distribution), Education, Income, and Ruling parties support. For each variable, the table reports the sample size, mean, and standard deviation for both the control and treatment groups. Additionally, an F -test statistic (and corresponding p -value) is provided to assess the statistical significance of differences between the groups. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

In Table D1, several covariates show notable differences between the control and treatment groups. For instance, the mean age is significantly higher in the treatment group (Mean = 63, SD = 17) compared to the control group (Mean = 50, SD = 14), with an F -statistic of 118.785, which is highly significant ($p < 0.01$). Similarly, the proportion of females is substantially higher in the treatment group (Mean = 0.69, SD = 0.46) than in the control group (Mean = 0.32, SD = 0.47), with an F -statistic of 88.224 ($p < 0.01$). These significant differences suggest that the treatment and control groups are not well-balanced in terms of age and gender, which could potentially bias the results in the case of the first gold medal. In contrast, ruling parties support does not show a significant difference between the groups ($F = 0.2$, $p > 0.1$).

In Table D2, a different pattern emerges. While age ($F = 5.528$, $p < 0.05$) and gender ($F = 14.238$, $p < 0.01$) are significantly different between the groups, other covariates like education ($F = 0.045$, $p > 0.1$), income ($F = 0.253$, $p > 0.1$), and ruling parties support (F

Treatment	Control (Before)			Treatment (After)			
Variable	N	Mean	SD	N	Mean	SD	Test
Age	226	51	14	486	54	17	$F = 5.528^{**}$
Female	226	0.3	0.46	486	0.45	0.5	$F = 14.238^{***}$
Education	226	3.7	6.4	486	3.8	7.6	$F = 0.045$
Income	226	7	3.4	486	6.8	3.1	$F = 0.253$
In-partisans	226	0.35	0.48	486	0.32	0.47	$F = 0.568$

Table D2: Balance check results also represent a clear imbalance between the treatment and control groups (The second gold medal case)

Note: See also *Note* in Table D1.

$= 0.568, p > 0.1$) do not show significant differences. This indicates that while some variables are imbalanced, others are relatively comparable between the control and treatment groups in the case of the second gold medal.

Therefore, our balance check results indicate significant imbalances in several covariates between the control and treatment groups, particularly in terms of age and gender, in both datasets. These imbalances could introduce confounding effects, making it essential to control for these variables in subsequent analyses to ensure that the observed treatment effects are not biased by these differences. To address this, we will re-assess the results through an entropy balancing and a counterfactual analysis.

Entropy balancing

We use entropy balancing to improve the covariate balance between the treatment and control groups, ensuring that the groups are comparable for estimating the causal effects of the Olympic events (Hainmueller, 2012). Entropy balancing reweights the control group to match the covariate distributions of the treatment group, creating a more balanced sample across key variables like age, income, and political attitudes. As a supplementary step to our main analysis, the key advantage of entropy balancing over traditional matching methods is that it guarantees exact balance on selected covariates without the need for iterative adjustments. This allows us to reduce model dependency and maintain the precision of our estimates, thereby presenting the results as supplementary evidence that enhances the robustness of our causal inferences.

In Tables D3—D6, we find covariate distributions both before and after entropy balancing for variables like age, education, income, ruling party support, and patriotism. The results demonstrate that entropy balancing effectively achieves balance across the covariates, evidenced by the alignment of the mean, variance, and skewness between the control and treatment groups after balancing. In Table D7—D8, we see regression analyses conducted with and without the entropy balancing weights. Despite achieving covariate balance, the treatment effect (medal effect) remains statistically insignificant, reinforcing the null hypothesis that the Olympic victories did not lead to any measurable increase in support for the incumbent government.

Treatment	Control (Before)			Treatment (After)		
Variable	Mean	Variance	Skewness	Mean	Variance	Skewness
Age	49.85	186.68	-0.07	63.40	291.09	-1.70
Education	3.29	0.83	-0.81	2.98	0.85	-0.10
Income	7.18	10.86	0.27	6.02	7.99	0.82
In-partisans	0.30	0.21	0.86	0.32	0.22	0.77
Patriotism	-0.03	0.85	-0.80	0.12	0.61	-0.82

Table D3: Covariate distribution *BEFORE* entropy balancing (The first gold medal case)

Note: This table presents the distribution of covariates *before* applying entropy balancing, specifically for the first gold medal case. The table compares the covariate distributions between the control group (Before) and the treatment group (After) using three statistics: Mean, Variance, and Skewness. These covariates include Age, Education, Income, Ruling parties support, and Patriotism.

Treatment	Control (Before)			Treatment (After)		
Variable	Mean	Variance	Skewness	Mean	Variance	Skewness
Age	53.18	184.66	-0.07	53.18	482.54	-1.70
Education	3.22	0.84	-0.81	3.22	0.87	-0.10
Income	6.90	10.41	0.27	6.90	11.60	0.82
In-partisans	0.31	0.21	0.86	0.31	0.21	0.77
Partriotism	0.05	0.84	-0.89	0.05	0.74	-0.91

Table D4: Covariate distribution *AFTER* entropy balancing (The first gold medal case)

Note: This table presents the distribution of covariates *after* applying entropy balancing, specifically for the first gold medal case. See also *Note* in Table D3.

Treatment	Control (Before)			Treatment (After)		
Variable	Mean	Variance	Skewness	Mean	Variance	Skewness
Age	50.77	183.90	-0.16	53.75	276.76	-0.36
Education	3.65	41.41	14.39	3.77	57.30	12.28
Income	6.97	11.53	0.36	6.84	9.65	0.43
In-partisans	0.35	0.23	0.63	0.32	0.22	0.76
Patriotism	-0.02	0.86	-0.71	0.01	0.84	-0.79

Table D5: Covariate distribution *BEFORE* entropy balancing (The second gold medal case)

Note: This table presents the distribution of covariates *before* applying entropy balancing, specifically for the second gold medal case. See also *Note* in Table D3.

Treatment	Control (Before)			Treatment (After)		
Variable	Mean	Variance	Skewness	Mean (Treatment)	Variance	Skewness
Age	52.80	177.85	-0.16	52.80	274.87	-0.36
Education	3.73	55.83	14.39	3.73	52.71	12.28
Income	6.88	11.33	0.36	6.88	9.79	0.43
In-partisans	0.33	0.22	0.63	0.33	0.22	0.76
Patriotism	-0.00	0.88	-0.71	0.00	0.84	-0.79

Table D6: Covariate distribution *AFTER* entropy balancing (The second gold medal case)

Note: This table presents the distribution of covariates *after* applying entropy balancing, specifically for the second gold medal case. See also *Note* in Table D3.

Table D7: Regression Results with and without Weights (The first gold medal case)

	<i>Dependent variable:</i>			
	<i>Cabinet approval</i>	<i>Feeling thermometer</i>	<i>Cabinet approval</i>	<i>Feeling thermometer</i>
	No Weight (1)	No Weight (2)	Weighted (3)	Weighted (4)
Treatment (before=0; After=1)	0.021 (0.037)	1.240 (2.281)	0.031 (0.031)	1.020 (1.967)
Age	-0.002* (0.001)	0.084 (0.059)	-0.002* (0.001)	0.075 (0.055)
Female	0.024 (0.030)	-1.921 (1.866)	0.012 (0.029)	-2.687 (1.802)
Education	0.013 (0.016)	0.654 (0.961)	0.009 (0.016)	0.453 (0.981)
Income	0.001 (0.004)	-0.023 (0.271)	0.002 (0.004)	-0.050 (0.273)
In-partisans	0.454*** (0.030)	24.078*** (1.843)	0.478*** (0.030)	26.079*** (1.870)
Patriotism	0.025* (0.015)	5.557*** (0.939)	0.026* (0.015)	4.937*** (0.951)
Constant	0.121 (0.076)	27.227*** (4.722)	0.125 (0.076)	28.289*** (4.796)
Weighted	NO	NO	YES	YES
Observations	740	740	740	740
R^2	0.258	0.252	0.282	0.256
Adjusted R^2	0.251	0.245	0.275	0.249
Residual Std. Error (df = 732)	0.368	22.714	0.365	23.040
F Statistic (df = 7; 732)	36.425***	35.178***	41.104***	35.962***

Note: This table presents the results of Ordinary Least Squares (OLS) regression analyses for two dependent variables: Cabinet approval and Feeling thermometer, in response to the occurrence of the *first* gold medal. The regressions are reported with and without weights derived from entropy balancing. Columns (1) and (3) report the results for Cabinet approval, and columns (2) and (4) report the results for Feeling thermometer. Robust standard errors are in parentheses. Significance levels are indicated by * $p < 0.1$, ** $p < 0.05$, and *** $p < 0.01$.

Table D8: Regression Results with and without Weights (The second gold medal case)

	<i>Dependent variable:</i>			
	<i>Cabinet Approval</i>	<i>Feeling Thermometer</i>	<i>Cabinet Approval</i>	<i>Feeling Thermometer</i>
	No Weight (1)	No Weight (2)	Weighted (3)	Weighted (4)
Treatment (Before=0; After=1)	0.072** (0.029)	0.502 (1.819)	0.067** (0.029)	0.441 (1.803)
Age	-0.0003 (0.001)	0.045 (0.055)	-0.0003 (0.001)	0.034 (0.055)
Female	-0.024 (0.028)	-0.683 (1.729)	-0.027 (0.027)	-0.751 (1.721)
Education	-0.0004 (0.002)	0.050 (0.116)	-0.001 (0.002)	0.097 (0.115)
Income	-0.004 (0.004)	-0.133 (0.264)	-0.005 (0.004)	-0.087 (0.265)
In-partisans	0.475*** (0.029)	24.235*** (1.795)	0.479*** (0.029)	24.240*** (1.807)
Patriotism	0.012 (0.015)	3.455*** (0.937)	0.010 (0.015)	3.647*** (0.935)
Constant	0.076 (0.064)	29.281*** (4.004)	0.090 (0.066)	29.472*** (4.133)
Weighted	NO	NO	YES	YES
Observations	712	712	712	712
R^2	0.290	0.230	0.294	0.231
Adjusted R^2	0.283	0.222	0.287	0.224
Residual Std. Error (df = 704)	0.355	22.270	0.355	22.396
F Statistic (df = 7; 704)	41.054***	30.066***	41.933***	30.241***

Note: See also Note in Table D7.

Counterfactual analysis with causal forest approach

To address the imbalance between the control and treatment groups, we implemented a causal forest algorithm (CFA) to generate counterfactual predictions, following the methodology outlined by Athey et al. (2019) and Wager and Athey (2018). Using the CFA, we estimated the conditional average treatment effect (CATE) by predicting what each respondent’s outcome—specifically, cabinet approval and feeling thermometer—would have been under different treatment scenarios (Athey and Wager, 2019). The model integrates treatment assignment, covariates, and the outcome variables to estimate the treatment effect for each individual by comparing the predicted outcomes across different treatment conditions.

Outcome variables	Results
Cabinet approval	0.017 [0.106, -0.077]
Feeling thermometer	-1.417 [6.330, -5.728]

Table D9: Counterfactual analysis with causal forest shows the null effect from event (The first gold medal)

Note: This table presents the results of a counterfactual analysis using a causal forest. The results are machine learning-based estimates for the scenario where each individual is assigned different covariate values. The results show the median (50th percentile) of the simulation outcomes, with the 2.5th and 97.5th percentiles in parentheses.

Outcome variables	Results
Cabinet approval	0.056 [0.138, -0.043]
Feeling thermometer	-0.385 [3.017, -3.445]

Table D10: Counterfactual analysis with causal forest also shows the null effect from event (The second gold medal)

Note: See also *Note* in Table D9.

The CFA results (Table D9—D10) did not reveal a statistically significant CATE for cabinet support or the feeling thermometer in either of the medal events. The analysis addressing the imbalance between the treatment and control groups upholds the findings of null effects for the Olympic events.

D.3 Visualizing Rolling RDD Results

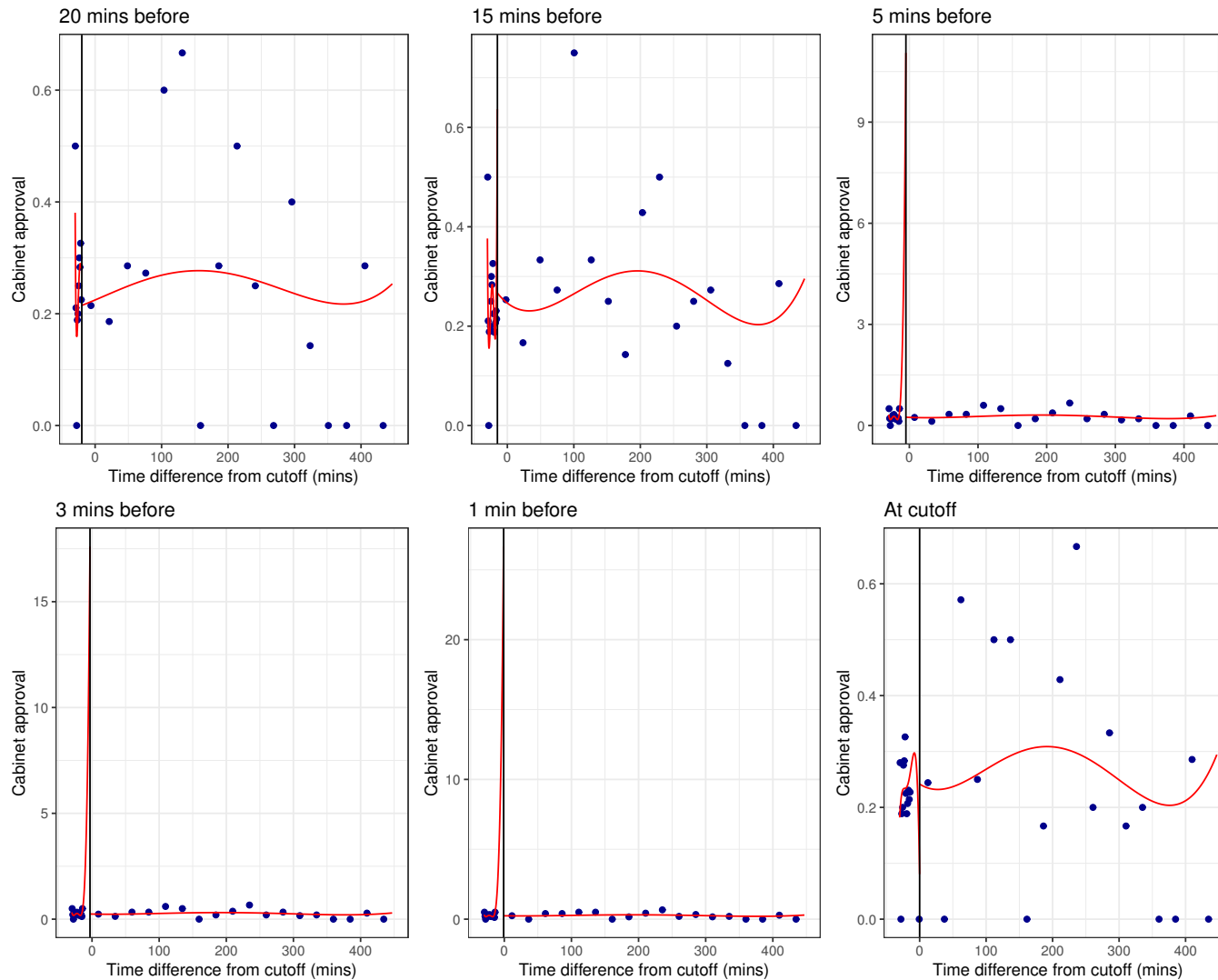


Figure D5: Gradually shifting the cutoff (*Before* the first gold medal) does not alter the null results for cabinet approval

Note: This figure presents the RDD result that we estimate the first gold medal effect using nonparametric kernel-based local polynomials on either side of the threshold. This figure also demonstrates the estimated RDDs with cutoff points shifted by -20, -15, -5, -3, and -1 minute. The cutoff is set at the gold medal confirmation. The outcome variables are cabinet approval and feeling thermometer to the Japanese government. The running variable is the time elapsed from the survey beginning to the survey end. The curve represents the fitted line from the polynomial regression results, and each point indicates the point estimate at each time point (every minute).

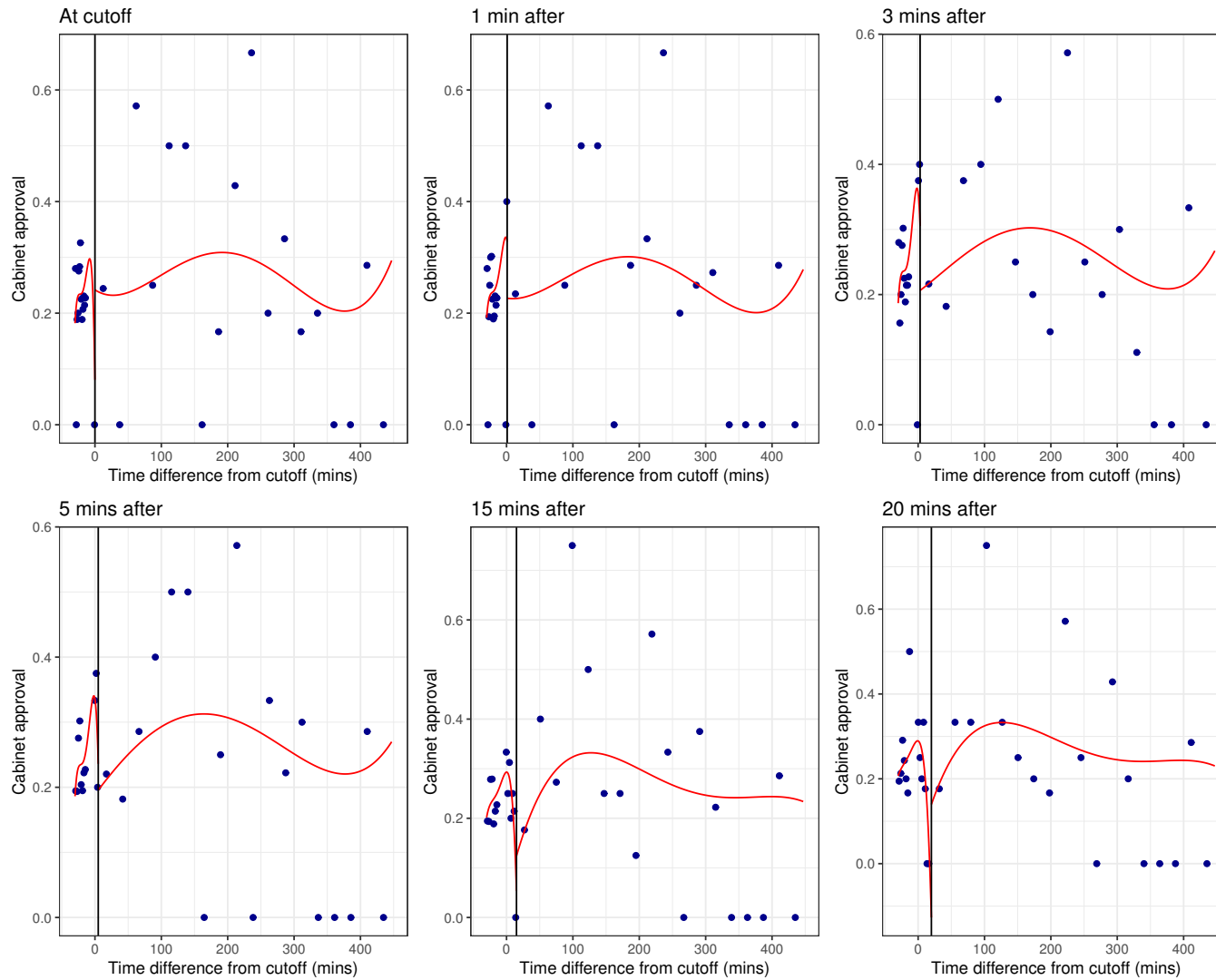


Figure D6: Gradually shifting the cutoff (*After* the first gold medal) does not alter the null results for cabinet approval

Note: This figure presents the RDD result that we estimate the first gold medal effect using nonparametric kernel-based local polynomials on either side of the threshold. This figure also demonstrates the estimated RDDs with cutoff points shifted by +20, +15, +5, +3, and +1 minute. See also Note in Figure D5.

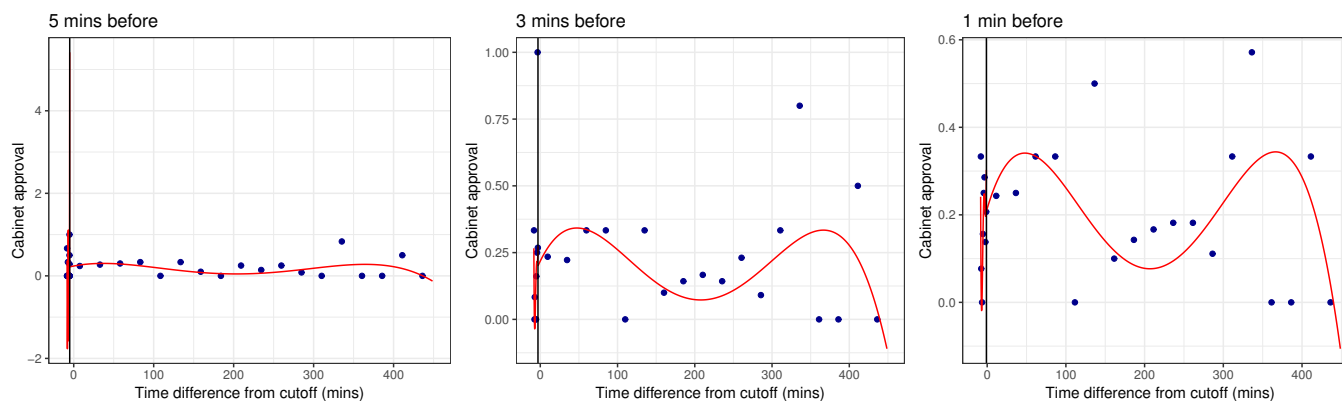


Figure D7: Gradually shifting the cutoff (*Before* the second gold medal) does not alter the null results for cabinet approval

Note: This figure presents the RDD result that we estimate the second gold medal effect using nonparametric kernel-based local polynomials on either side of the threshold. See also *Note* in Figure D5.

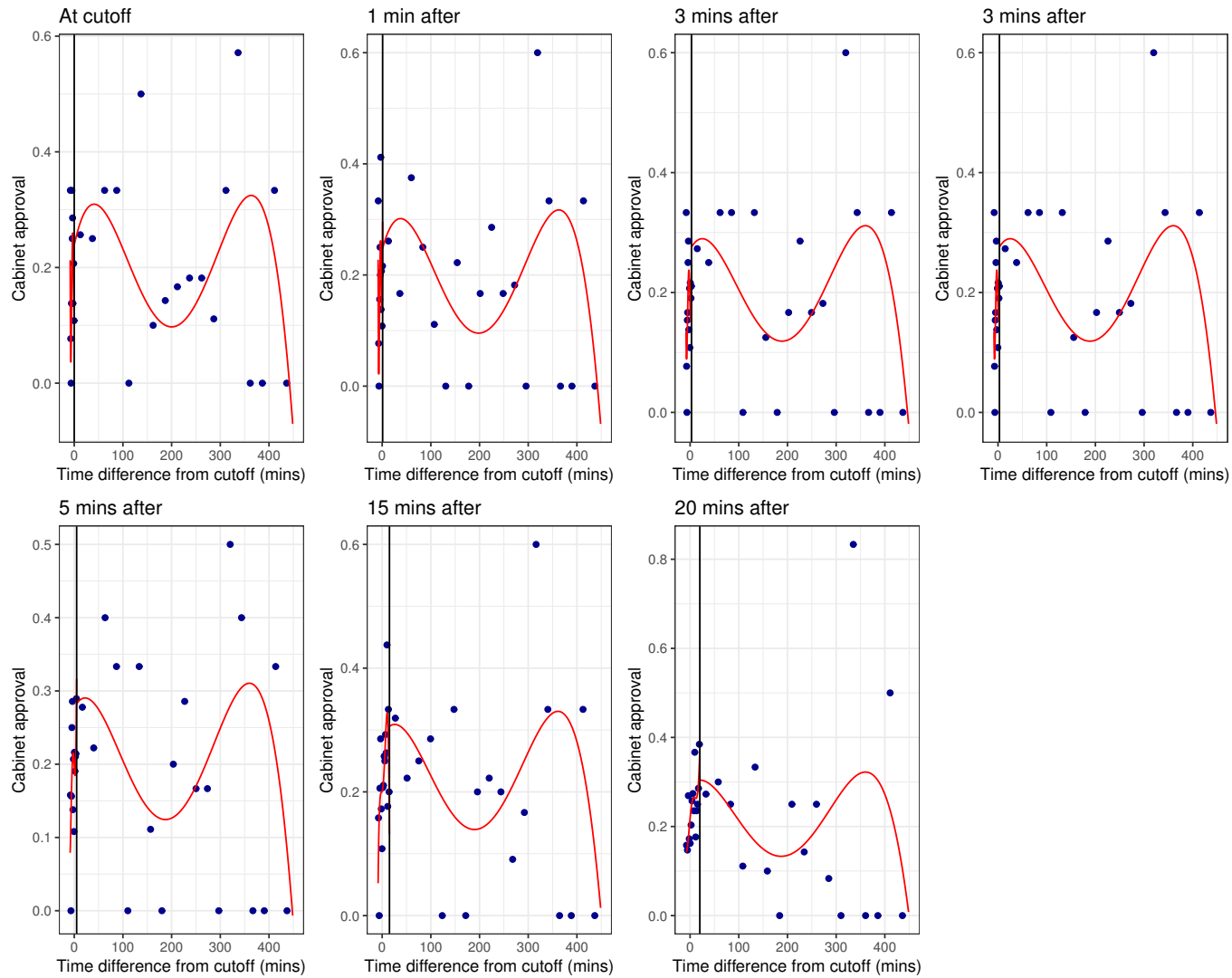


Figure D8: Gradually shifting the cutoff (*After* the second gold medal) does not alter the null results for cabinet approval

Note: This figure presents the RDD result that we estimate the second gold medal effect using nonparametric kernel-based local polynomials on either side of the threshold. See also *Note* in Figure D6.

D.4 No response analysis

If there is a systematic bias in the attrition of the outcome variable between the treatment and control groups, it could lead to a biased estimation of the treatment effect size. To address this concern, we followed Muñoz et al. (2020)’s recommendations and closely examined the no-response rate between the treatment and control groups. Table D11 reveal no statistically significant differences in the outcome variable or in the no-response status to debriefing, which indicates the final completion of the survey, in both cases.

It is also crucial to assess whether there is any bias in non-response rates across different attributes among the groups. Given that there is no significant bias in non-response by attribute across the groups, the risk of estimator distortion due to attrition-related excludability violations can be considered minimal.

Variable	Control	Treatment	<i>t</i> -statistics	<i>p</i> -value
<i>The first gold medal case</i>				
Age	0.09	0.12	-1.70	0.09
Female	0.09	0.12	-1.72	0.09
Education	0.09	0.12	-1.78	0.08
Income	0.21	0.24	-1.12	0.26
In-partisans	0.10	0.14	-1.62	0.11
<i>The second gold medal case</i>				
Age	0.05	0.04	0.74	0.46
Female	0.06	0.05	0.84	0.40
Education	0.05	0.04	0.74	0.46
Income	0.18	0.19	-0.41	0.68
In-partisans	0.10	0.12	-1.15	0.25

Table D11: *t*-test results indicate no major difference of non-response rate across covariates

Note: This table presents the results of *t*-tests comparing the non-response rates between the control and treatment groups for two different cases: the first gold medal and the second gold medal. The table is divided into two sections, each corresponding to one of the gold medal cases. For each variable (Age, Female, Education, Income, and Ruling parties support), the table shows the mean non-response rates in the control and treatment groups, along with the *t*-statistics and the associated *p*-values.

D.5 Time trends

If there is a prior trend in the outcome variable within the control group, the effect of the event occurrence may be overestimated and biased toward upward. Following Muñoz et al. (2020)’s recommendation, we conducted an analysis to examine the time trends before (and after) the event.

First, to assess temporal stability, as suggested by Muñoz et al. (2020), we restricted our sample to the pre-event data and examined the placebo treatment effect by simulating the event at the pre-event median, as shown in Table D12—D15. Column 1 results in Table D14 and D15 demonstrate that the interaction terms of placebo treatment and times show only minor, statistically significant changes in cabinet support and the feeling thermometer. The treatment effects indicated by these results are negative and do not align with theoretical predictions. In the other models, both the treatment and the time course results show no significant effects, with no evidence of any trend existing prior to the treatment.

	<i>Dependent variable:</i>			
	<i>Cabinet approval</i>			
	(1)	(2)	(3)	(4)
Placebo treatment	−0.001 (0.075)	−0.007 (0.074)		0.022 (0.073)
Times	0.023 (0.019)	0.016 (0.018)	0.0005 (0.005)	−0.002 (0.010)
Times ²	0.001 (0.001)		−0.001 (0.001)	−0.001 (0.001)
Placebo treatment $\times$ times	−0.044 (0.028)	−0.028 (0.020)		
Constant	0.287*** (0.064)	0.278*** (0.064)	0.242*** (0.020)	0.228*** (0.047)
Observations	558	558	558	558
R^2	0.006	0.005	0.001	0.002
Adjusted R^2	−0.002	−0.001	−0.002	−0.004
Residual Std. Error	0.423 (df = 553)	0.423 (df = 554)	0.424 (df = 555)	0.424 (df = 554)
F Statistic	0.775 (df = 4; 553)	0.851 (df = 3; 554)	0.382 (df = 2; 555)	0.290 (df = 3; 554)

Table D12: The falsification test using the placebo treatment (the median cutoff point of the control group) also detects a null effect on cabinet approval (the first gold medal case)

Note: This table presents the OLS estimation results using the sample positioned to the left of the cutoff, with the median point within the control group used as the placebo treatment. *: $p < 0.1$; **: $p < 0.05$; ***: $p < 0.01$. Values in parentheses represent robust standard errors.

	<i>Dependent variable:</i>			
	<i>Feeling thermometer to Japanese government</i>			
	(1)	(2)	(3)	(4)
Placebo treatment	0.209 (4.499)	0.431 (4.458)		0.506 (4.311)
Times	0.096 (1.166)	0.349 (1.113)	−0.158 (0.330)	−0.219 (0.611)
Times ²	−0.037 (0.048)		−0.058* (0.033)	−0.055 (0.041)
Placebo treatment×Times	−0.556 (1.618)	−1.184 (1.234)		
Constant	41.445*** (3.950)	41.779*** (3.919)	41.015*** (1.212)	40.709*** (2.939)
Observations	558	558	558	558
R^2	0.005	0.005	0.005	0.005
Adjusted R^2	−0.002	−0.001	0.001	−0.001
Residual Std. Error	26.624 (df = 553)	26.605 (df = 554)	26.578 (df = 555)	26.602 (df = 554)
F Statistic	0.682 (df = 4; 553)	0.844 (df = 3; 554)	1.316 (df = 2; 555)	0.881 (df = 3; 554)

Table D13: The falsification test using the placebo treatment (the median cutoff point of the control group) also detects a null effect on feeling thermometer (the first gold medal case)

Note: See also *Note* in Table D12.

We also applied the Augmented Dickey-Fuller test and fractional integration analysis to determine whether time trends existed before and after the treatment (Table D16). The results of the Augmented Dickey-Fuller test allow us to reject the null hypothesis that the series has a unit root before and after the treatment for all outcome variables. Additionally, the fractional integration analysis indicates that the series is stationary if $d < 0.5$, and each d value satisfies this condition. Therefore, we conclude that while there is a time trend before and after the treatment, the likelihood of this trend significantly influencing the results is minimal.

	<i>Dependent variable:</i>			
	<i>Cabinet approval</i>			
	(1)	(2)	(3)	(4)
Placebo treatment	−0.121 (0.147)	0.028 (0.106)		0.068 (0.102)
Times	0.193 (0.128)	0.020 (0.035)	−0.011 (0.012)	−0.024 (0.022)
Times ²	0.031 (0.023)		−0.006 (0.005)	−0.006 (0.005)
Placebo treatment×Times	−0.335* (0.201)	−0.070 (0.045)		
Constant	0.408*** (0.155)	0.234*** (0.089)	0.220*** (0.037)	0.179** (0.072)
Observations	226	226	226	226
R ²	0.024	0.015	0.009	0.011
Adjusted R ²	0.006	0.002	−0.0003	−0.002
Residual Std. Error	0.392 (df = 221)	0.393 (df = 222)	0.393 (df = 223)	0.394 (df = 222)
F Statistic	1.361 (df = 4; 221)	1.160 (df = 3; 222)	0.966 (df = 2; 223)	0.819 (df = 3; 222)

Table D14: The falsification test using the placebo treatment (the median cutoff point of the control group) also detects a null effect on cabinet approval (the second gold medal case)

Note: See also *Note* in Table D12.

	<i>Dependent variable:</i>			
	<i>Feeling thermometer to Japanese government</i>			
	(1)	(2)	(3)	(4)
Placebo treatment	−11.546 (9.439)	4.494 (5.961)		3.583 (5.711)
Times	16.103** (8.183)	−2.412 (1.972)	−0.537 (0.777)	−1.211 (1.299)
Times ²	3.360** (1.457)		0.353 (0.305)	0.356 (0.304)
Placebo treatment×Times	−26.793** (12.407)	1.690 (2.603)		
Constant	53.675*** (9.501)	34.951*** (5.005)	37.482*** (2.191)	35.328*** (4.065)
Observations	226	226	226	226
R ²	0.033	0.008	0.010	0.012
Adjusted R ²	0.016	−0.006	0.001	−0.001
Residual Std. Error	24.463 (df = 221)	24.725 (df = 222)	24.641 (df = 223)	24.675 (df = 222)
F Statistic	1.896 (df = 4; 221)	0.587 (df = 3; 222)	1.152 (df = 2; 223)	0.891 (df = 3; 222)

Table D15: The falsification test using the placebo treatment (the median cutoff point of the control group) also detects a null effect on feeling thermometer (the second gold medal case)

Note: See also *Note* in Table D12.

Test	Dickey-Fuller & FI d -value	Lag.Order	p -value for ADF test & SD for FI
<i>The first gold medal case</i>			
ADF Test (cabap)	-8.38	8.00	0.01
ADF Test (ftj)	-8.53	8.00	0.01
fdGPH (cabap)	-0.06	0.17	0.18
fdGPH (ftj)	-0.26	0.17	0.19
<i>The second gold medal case</i>			
ADF Test (cabap)	-6.05	6.00	0.01
ADF Test (ftj)	-4.43	6.00	0.01
fdGPH (cabap)	0.03	0.22	0.18
fdGPH (ftj)	0.10	0.22	0.21

Table D16: Dickey-Fuller Test and Fractional Integration d -value suggest both before and after series are stationary and have no trends

Note: This table presents the results of the augmented Dickey-Fuller (ADF) test and the fractional integration (FI) estimation. The null hypothesis of the ADF test is that the series has a unit root, meaning it is non-stationary. The FI test determines that the series is stationary when the d -value is less than 0.5.

References

- Achen, Christopher H.. and Bartels, Larry M.. (2004), ‘Blind Retrospection: Electoral Responses to Drought, Flu, and Shark Attacks’, *Research Concentration: Elections and Electoral Rules* **5**.
- Achen, Christopher H.. and Bartels, Larry M.. (2016), *Democracy for Realists : Why Elections Do not Produce Responsive Government*, Princeton University Press, Princeton, NJ.
- Achen, Christopher H., Bartels, Larry M., Achen, Christopher H.. and Bartels, Larry M.. (2018), *Democracy for Realists*, Princeton University Press.
- Athey, Susan., Tibshirani, Julie. and Wager, Stefan. (2019), ‘Generalized random forests’, *Ann. Statist.* **47**(2), 1148–1178.
- Athey, Susan. and Wager, Stefan. (2019), ‘Estimating Treatment Effects with Causal Forests: An Application’, *Observational Studies* **5**(2), 37–51.
- Catalinac, Amy. and Muraoka, Taishi. (2023), ‘How incumbent politicians respond to the enactment of a programmatic policy: evidence from snow subsidies’, *Japanese Journal of Political Science* **24**(1), 41–63.
- Chang, Chun Ping. and Berdiev, Aziz N.. (2015), ‘Do natural disasters increase the likelihood that a government is replaced?’, *Applied Economics* **47**(17), 1788–1808.
- Cole, Shawn., Healy, Andrew. and Werker, Eric. (2012), ‘Do voters demand responsive governments? Evidence from Indian disaster relief’, *Journal of Development Economics* **97**(2), 167–181.
- Cooperman, Alicia. (2022), ‘(Un)Natural Disasters: Electoral Cycles in Disaster Relief’, *Comparative Political Studies* **55**(7), 1158–1197.
- Fowler, Anthony. and Hall, Andrew B.. (2018), ‘Do shark attacks influence presidential elections? Reassessing a prominent finding on voter competence’, *Journal of Politics* **80**(4), 1423–1437.
- Fowler, Anthony. and Pablo Montagnes, B.. (2015), ‘College football, elections, and false-positive results in observational research’, *Proceedings of the National Academy of Sciences of the United States of America* **112**(45), 13800–13804.
- Fukumoto, Kentaro. and Kikuta, Kyosuke. (2024), ‘After a Storm Come Votes: Identifying the Effects of Disaster Relief on Electoral Outcomes’, *Political Behavior* pp. 1–21.
- Gallego, Jorge. (2018), ‘Natural disasters and clientelism: The case of floods and landslides in Colombia’, *Electoral Studies* **55**, 73–88.

- Gasper, John T.. and Reeves, Andrew. (2011), ‘Make It Rain? Retrospection and the Attentive Electorate in the Context of Natural Disasters’, *American Journal of Political Science* **55**(2), 340–355.
- Graham, Matthew H., Huber, Gregory A., Malhotra, Neil. and Mo, Cecilia Hyunjung. (2023), ‘Irrelevant Events and Voting Behavior: Replications Using Principles from Open Science’, *Journal of Politics* **85**(1), 296–303.
- Hainmueller, Jens. (2012), ‘Entropy Balancing for Causal Effects: A Multivariate Reweighting Method to Produce Balanced Samples in Observational Studies’, *Political Analysis* **20**(1), 25–46.
- Healy, Andrew J., Malhotra, Neil. and Mo, Cecilia Hyunjung. (2010), ‘Irrelevant events affect voters’ evaluations of government performance’, *Proceedings of the National Academy of Sciences of the United States of America* **107**(29), 12804–12809.
- Hilbig, Hanno. and Riaz, Davis Sascha. (2024), ‘Natural Disasters and Green Party Support’, *Journal of Politics* **86**(1), 241–256.
- Kondoh, Haruo. and Miyazaki, Takeshi. (2024), ‘Do natural disasters affect local elections? An empirical examination using subnational electoral data’, *Economics & Politics* **36**(2), 868–900.
- Masiero, Giuliano. and Santarossa, Michael. (2021), ‘Natural disasters and electoral outcomes’, *European Journal of Political Economy* **67**, 101983.
- Müller, Stefan. and Kneafsey, Liam. (2023), ‘Evidence for the irrelevance of irrelevant events’, *Political Science Research and Methods* **11**(2), 311–327.
- Muñoz, Jordi., Falcó-Gimeno, Albert. and Hernández, Enrique. (2020), ‘Unexpected Event during Survey Design: Promise and Pitfalls for Causal Inference’, *Political Analysis* **28**(2), 186–206.
- Rapeli, Lauri. and Söderlund, Peter. (2022), ‘Does sports success increase government support? Voter (ir)rationality in a multiparty context’, *Research and Politics* **9**(3).
- Visconti, Giancarlo. (2022), ‘After the Flood: Disasters, Ideological Voting and Electoral Choices in Chile’, *Political Behavior* **44**(4), 1985–2004.
- Wager, Stefan. and Athey, Susan. (2018), ‘Estimation and Inference of Heterogeneous Treatment Effects using Random Forests’, *Journal of the American Statistical Association* **113**(523), 1228–1242.