Randomization Inference for Before-and-After Studies with Multiple Units: An Application to a Criminal Procedure Reform in Uruguay

Supplemental Appendix

Matias D. Cattaneo[*] Carlos Díaz[†] Rocío Titiunik[‡]

August 8, 2025

[*]Department of Operations Research and Financial Engineering, Princeton University.

[†]Millennium Nucleus on Criminal Complexity, and School of Economics and Business, Universidad Alberto Hurtado.

[‡]Department of Politics, Princeton University.

Contents

SA-1 Results with Time-Adjusted Test Statistic

In the main paper, we presented Fisherian randomization results using the unadjusted difference-in-means as the test statistic. We now present results using an alternative test statistic that removes the time trends in the outcome using a linear specification. We fit a linear model of the outcome on a unit-specific intercept and a unit-specific linear time trend, using a window of 300 hundred days before and after the adoption of the real intervention for each unit. The time-adjusted test statistic is the difference in the average residuals of this fit between treatment and control groups. This effectively produces a de-trended version of the outcome. The interpretation of these results under our framework requires that we invoke Assumptions 1 and 2 for the adjusted outcomes. The results are presented in Table SA-1, employing the TR assignment mechanism.

Table SA-1: Short-Term Effects of CCP Reform in Montevideo
Outcome: Daily number of crimes reported to police
Actual adoption time: November 1, 2017

	Estimates		TR Mechanism	
τ	$\widehat{\theta}_\tau$	$\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i,\tau,0}$	p-value	95% CI
1 day	0.835	0.006	0.131	[-0.22, 1.90]
7 days	0.375	-0.397	0.021	[0.06, 0.69]
14 days	0.243	-0.210	0.055	[0.00, 0.49]

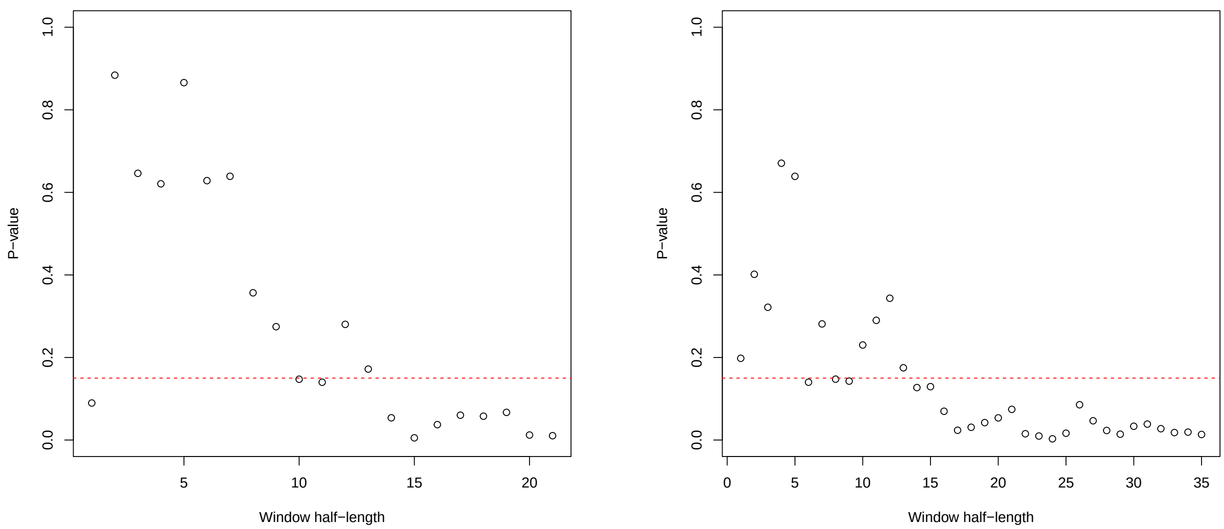
Sample is 62 neighborhoods in Montevideo, each observed before and after the adoption of the CCP reform. The p-values are randomization-based for $\widehat{\theta}_\tau = \frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i,\tau,1} - \frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i,\tau,0}$ with 10,000 simulations. τ denotes the half-length of a symmetric window around adoption time. CI refers to confidence interval, calculated by inversion based on a constant treatment effect model. TR refers to Treatment Reversal mechanism.

The advantage of using time-adjusted outcomes is that it potentially allows researchers to consider larger windows, as the assumption of no trends in the TR mechanism is more plausible for the de-trended outcome. The disadvantage is that it requires the assumption that the functional form used to model the outcome trend is a good enough approximation. In real applications, it may not be easy to assess how plausible a given adjustment model is. Nonetheless, the extension to time-adjusted test statistics makes our framework more general and will be relevant in some applications.

SA-2 Additional Window Selection Results

We present the window selector described in the paper using the TR assignment mechanism, with two additional artificial adoption times: $a = -21$ (21 days before the actual adoption), and $a = -35$ (35 days before the actual adoption). The x-axis contains τ, the half-length of the window $\mathcal{W}_\tau$, while the y-axis shows the p-value associated with a test of the Fisherian sharp null hypothesis that the treatment effect in the corresponding window $\mathcal{W}_\tau$ is zero for all units, using the TR assignment mechanism for the total crime outcome. The horizontal line is drawn at 0.15.

Figure SA-1: Window Selector Around *Placebo* Adoption Times



(a) 21 days before adoption

(b) 35 days before adoption

SA-3 Additional Falsification Results

We report additional falsification results. Table 4 in the main paper presents tests of $\mathsf{H}_{\mathtt{TR}}$ and $\mathsf{H}_{\mathtt{AT}}$ with the adoption time artificially set to midnight on November 1st for the years 2015, 2016, and 2018. We show that these results remain robust when we consider alternative values of τ.

Figure SA-5 shows the average differences, $\widehat{\theta}_\tau$, and the randomization-based p-values for all values of τ between 1 and 21 (not just 1, 7, and 14, as reported in Table 4 in the main paper) around the different placebo adoption times considered. When we use the actual adoption time, the null hypothesis rejected at 5% level (grey bars are p-values) or below for almost all values of τ greater than 5, and the average differences between treated and untreated observed outcomes (red dots) are always positive and well above the values recorded for the same period in each of the placebo adoption times. In contrast, in the analyses that use placebo adoption times, p-values are generally above 10% and, in several cases, the effects are negative (note that the red dots show only positive estimates; negative estimates would be below the horizontal axis and are omitted).

We repeat these analyses, but instead of setting the artificial adoption time to November 1st, we set it to the same day of the week in different years. Considering that November 1st, 2017, was a Wednesday, we set the artificial adoption times to the first Wednesday of November in 2015, 2016, and 2018. The results are reported in Table SA-2, which is analogous to Table 4 in the main paper. Similarly to Table 4 in the main paper, the tests in Table SA-2 fail to reject the null hypotheses of no effects. Figure SA-6 is analogous to Figure SA-5, showing falsification results for all values of $\tau \geq 21$.

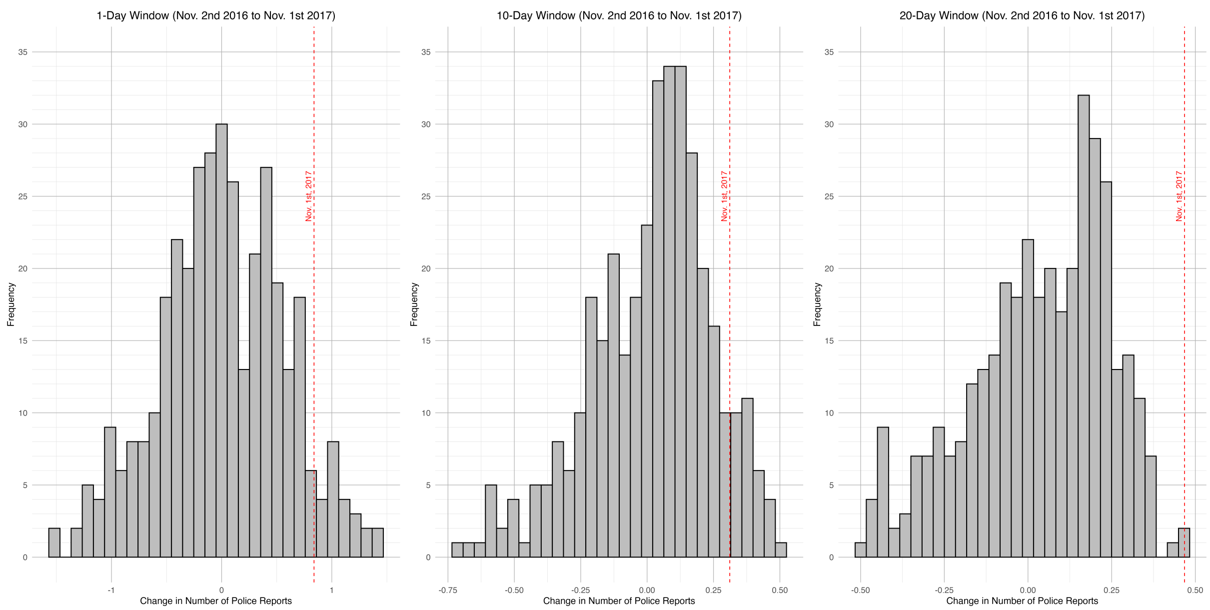
Table SA-2: Short-term Effects of CCP Reform in Montevideo
Around *Placebo* Adoption Times (Day)
Outcome: daily number of crimes reported to police
Actual adoption time: November 1st, 2017

	2015			2016			2018		
		P-Value			P-Value			P-Value	
τ	$\widehat{\theta}_\tau$	TR	AT	$\widehat{\theta}_\tau$	TR	AT	$\widehat{\theta}_\tau$	TR	AT
1 day	-0.194	0.689	-	-0.065	0.910	-	0.129	0.791	-
7 days	-0.106	0.515	0.513	0.046	0.772	0.918	-0.081	0.680	0.864
14 days	-0.015	0.914	0.879	0.142	0.230	0.809	0.046	0.714	0.744

Sample is 62 neighborhoods in Montevideo, each observed before and after the adoption of the CCP reform. The p-values are randomization-based for $\widehat{\theta}_\tau = \frac{1}{n}\sum_{i=1}^{n} \bar{Y}_{i,\tau,1} - \frac{1}{n}\sum_{i=1}^{n} \bar{Y}_{i,\tau,0}$ with 10,000 simulations. τ denotes the half-length of a symmetric window around adoption time. TR and AT refer to Treatment Reversal and Adoption Timing mechanisms, respectively. In AT mechanism, $\mathcal{A} = \{-6, -5, \ldots, 0\}$ for $\tau = 7$ and $\mathcal{A} = \{-13, -12, \ldots, 0\}$ for $\tau = 14$.

We also report, in Figure SA-8 the results from Table 4 for the first 21 days of November in the years 2017, 2018, and 2019. As shown, while there is an increase on the first day of the month in both 2018 and 2019, 2017 remains the only year in which effects emerge and persist during the early days of November, under both the TR and AT assignment mechanisms. In fact, 2017 is the only case that consistently exhibits non-negative effects across all window lengths—from 1 to 21 days.

Figure SA-2: Daily Change at Adoption Time vs. Prior-Year Daily Changes
(1-, 10-, and 20-Day Windows)



Aditionally, Figure SA-2 shows that the change on November 1, 2017, is exceptionally large relative to the distribution of daily changes across all 1-, 10-, and 20-day windows in the year preceding the introduction of the new CCP (November 2, 2016, to November 1, 2017).

SA-4 Additional Results Assessing Potential Mechanisms

As we explained in the main paper, the reform was widely known to the public. The Office of the Attorney General launched a public information campaign informing the public about the reform. As a result, the reform received considerable media coverage. Figure SA-3 shows a billboard placed at several Montevideo bus stops highlighting the transparency and guarantees of the new procedural law.

Figure SA-3: Billboard Promoting CCP Reform

As argued in Section 4.2 in the main paper, we have found no evidence that the increases in crime reports that we see were caused by changes in how crimes are reported rather actual crimes.

The reform may have changed the costs victims face when reporting crimes. Although the new CCP allows individuals to report offenses directly at a prosecutor's office, our data show that nearly all crimes continued to be reported to the national police. Nonetheless, the reform may have coincided with a shift in police behavior (Hausman and Kronick, 2023) that could help explain our findings. By the time the reform took effect, officers were already equipped with tablets that enabled them to file crime reports at the scene, eliminating the need for victims to travel to a police station. If the observed increase in reports was driven by enhanced street-level policing, we would expect a concentration of reports shortly after the incidents occurred. As shown in Figure SA-7, we find no evidence that the effects we see are driven by reports filed within 15, 30, or 45 minutes of the event. This makes it unlikely that improved on-the-spot reporting alone can account for the increase we document.

We also compare the immediate impact of the new CCP on the number of reported thefts to its impact on the number of reported robberies (i.e., violent thefts) and domestic violence (the most frequent crime against persons). If the increase in crime reports is due to a more selective use of preventive detention, we should not observe any impact on domestic violence (a crime that is considered very severe by prosecutors under both procedural regimes), while the effects should be present for thefts and robberies—and should be higher for thefts than for robberies, as thefts are classified as non-violent and thus might avoid preventive detention.

Table SA-3 presents results. As expected, the number of domestic violence incidents reported to the police does not exhibit an immediate response to the change in CCP. In fact, negative effects are observed for $\tau = 1$ and $\tau = 7$ (null hypothesis not rejected). The effects for robberies also appear weak, with a decrease in the smallest window. Thefts is the only category that shows a consistent increase in all windows considered, although the p-values are higher than for the total effect, particularly for the AT mechanism.

This pattern suggests that the increase in the total number of police reports is not driven by the two most frequent violent crimes. In contrast, the results show that the reform might have increased thefts immediately after implementation. However, these effects are not as strong as the effects for total crime reported in Table 2, which suggests that the decrease in preventive detention is likely only part of the explanation.

Table SA-3: Short-Term Effects of CCP Reform in Montevideo
for Different Types of Crime

Outcome: daily number of crimes reported to police

Actual adoption time: November 1st, 2017

Property Crime: *Theft*						
	Estimates		TR Mechanism		AT Mechanism	
τ	$\widehat{\theta}_\tau$	$\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i,\tau,0}$	p-value	95% CI	p-value	95% CI
1 day	0.500	2.290	0.115	[-0.09, 1.10]	-	-
7 days	0.194	2.270	0.067	[0.00, 0.39]	0.338	[-0.40, 0.71]
14 days	0.150	2.363	0.162	[-0.05, 0.35]	0.185	[-0.04, 1.27]
Violent Crime: *Robbery*						
	Estimates		TR Mechanism		AT Mechanism	
τ	$\widehat{\theta}_\tau$	$\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i,\tau,0}$	p-value	95% CI	p-value	95% CI
1 day	-0.258	0.935	0.207	[-0.63, 0.11]	-	-
7 days	0.097	0.707	0.088	[-0.01, 0.20]	0.700	[-0.35, 0.19]
14 days	0.021	0.719	0.676	[-0.07, 0.11]	0.650	[-0.20, 0.22]
Violent Crime: *Domestic Violence*						
	Estimates		TR Mechanism		AT Mechanism	
τ	$\widehat{\theta}_\tau$	$\frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i,\tau,0}$	p-value	95% CI	p-value	95% CI
1 day	-0.097	0.613	0.560	[-0.37, 0.17]	-	-
7 days	-0.035	0.558	0.494	[-0.12, 0.05]	0.420	[-0.35, 0.14]
14 days	0.020	0.533	0.571	[-0.04, 0.08]	0.751	[-0.24, 0.19]

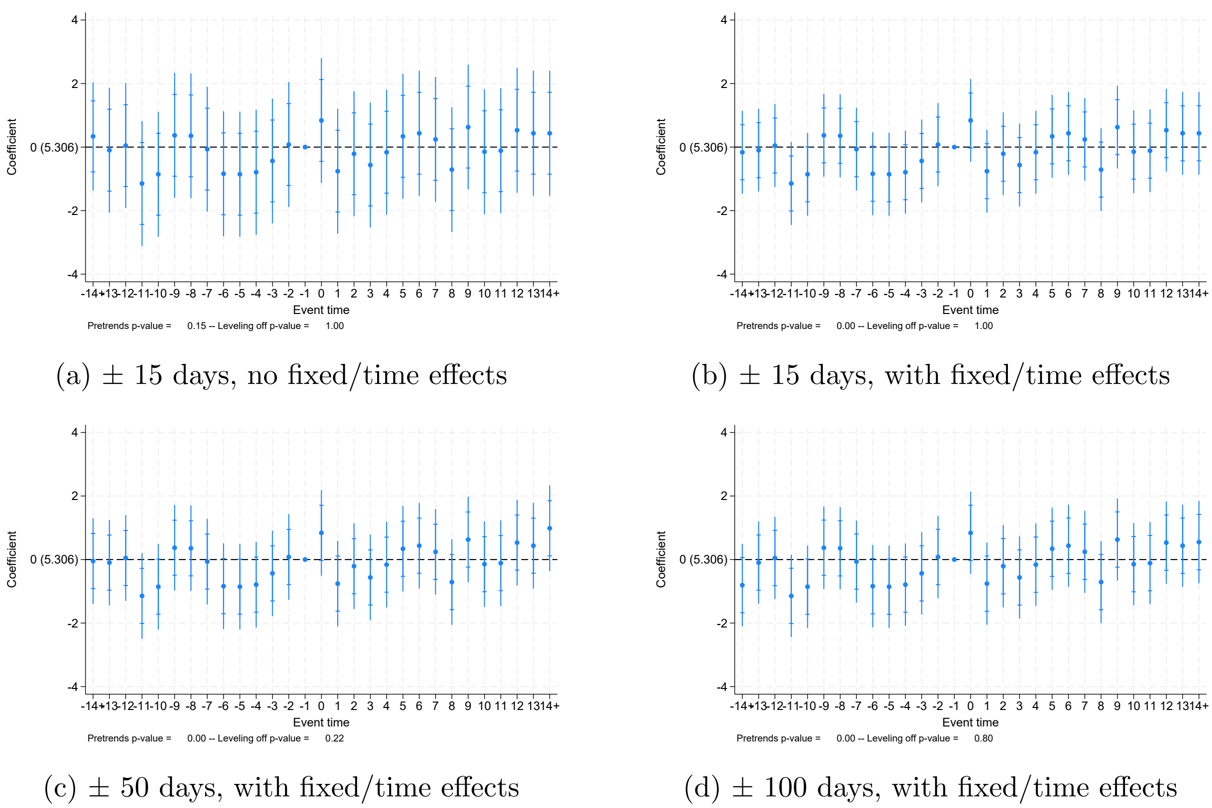
Sample is 62 neighborhoods in Montevideo, each observed before and after the adoption of the CCP reform. The p-values are randomization-based for $\widehat{\theta}_\tau = \frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i,\tau,1} - \frac{1}{n}\sum_{i=1}^{n}\bar{Y}_{i,\tau,0}$ with 10,000 simulations. τ denotes the half-length of a symmetric window around adoption time. CI refers to confidence interval, calculated by inversion based on constant treatment effect model. TR and AT refer to Treatment Reversal and Adoption Timing mechanisms, respectively. In AT mechanism, $\mathcal{A} = \{-6, -5, \dots, 0\}$ for $\tau = 7$ and $\mathcal{A} = \{-13, -12, \dots, 0\}$ for $\tau = 14$.

SA-5 Results Using Event Study Methods

We estimate effects using event study methods based on linear models, using $i = 1, \ldots n$ units and $t = 1, \ldots, T$ time periods. For implementation, we use the approach by Freyaldenhoven et al. (2019, 2025a), and fit a linear model of the outcome (q_{it}) on a unit fixed effect (α_i), a time fixed effect (γ_t), and binary policy variables that indicate whether each unit is treated in specific periods ($\sum_{j=-G}^{M} \beta_j z_{i,t-j}$), where the parameters β_j capture the effects of the policy. We created event study plots based on this fit, using the `xtevent` command (Freyaldenhoven et al., 2025b).

We fit this model in four different ways: (a) using 15 days before and 15 days after adoption, and excluding fixed-effects and time-effects; (b) using 15 days before and 15 days after adoption, including fixed-effects and time-effects; (c) using 50 days before and 50 days after adoption, including fixed-effects and time-effects; and (d) using 100 days before and 100 days after adoption, including fixed-effects and time-effects. The event-study plots are shown in Figure SA-4, where the outcome at the time period immediately before adoption is normalized to zero.

Figure SA-4: Event Study Plots for CCP Effect



(a) ± 15 days, no fixed/time effects

(b) ± 15 days, with fixed/time effects

(c) ± 50 days, with fixed/time effects

(d) ± 100 days, with fixed/time effects

In Figure SA-4(a), we exclude time effects and fixed effects, as both types of effects would

Table SA-4: Test that Cumulative CCP Effect at Time 0, 7 and 14 Are Jointly Zero

Model	F	`Prob` $> F$
(a)	0.58	0.6268
(b)	1.29	0.2766
(c)	2.26	0.0792
(d)	1.35	0.2566

be estimated with relatively few observations (only 30 periods for fixed effects, only 62 units for time effects). Overall, the cumulative effects show null effects between day -14 and day 14 after adoption of the intervention. Judging from these results, one would conclude that the short-run effects of the policy are zero. Adding fixed effects and time effects and keeping the number of days at 15, as shown in Figure SA-4(b), or adding additional days, as shown in Figure SA-4(c) and SA-4(d), does not change the overall pattern.

Moreover, for each model, we conducted a test that the coefficients capturing the cumulative effect of the policy at time 0, 7, and 14 are jointly zero using an F-test. Table SA-4 shows that the results vary considerably depending on how the model is specified. When only 15 days are used before/after adoption, the null hypothesis that the effects are jointly zero at time 0, 7 and 14 cannot be rejected, with very high p-value above 0.60. The same conclusion is drawn from model (b), where fixed effects and time effects are included. However, when the number of periods used to fit the linear model is increased to 50 days before/after adoption in model (c), the p-value drops to below 0.08. The p-value increases again for model (d). This illustrates how results can be sensitive to the parametric specification that is used to fit the model.

References

Freyaldenhoven, S., Hansen, C., Pérez, J. P., and Shapiro, J. M. (2025a), "Visualization, identification, and estimation in the linear panel event-study design," *Advances in Economics and Econometrics: Twelfth World Congress*.

Freyaldenhoven, S., Hansen, C., and Shapiro, J. M. (2019), "Pre-event trends in the panel event-study design," *American Economic Review*, 109, 3307–3338.

Freyaldenhoven, S., Hansen, C. B., Pérez, J. P., Shapiro, J. M., and Carreto, C. (2025b), "xtevent: Estimation and visualization in the linear panel event-study design," *The Stata Journal*, 25, 97–135.

Hausman, D., and Kronick, D. (2023), "The illusory end of stop and frisk in Chicago?" *Science advances*, 9, eadh3017.

Figure SA-5: Effects of CCP Reform for Different Windows Around *Placebo* Adoption Times (*Date*)

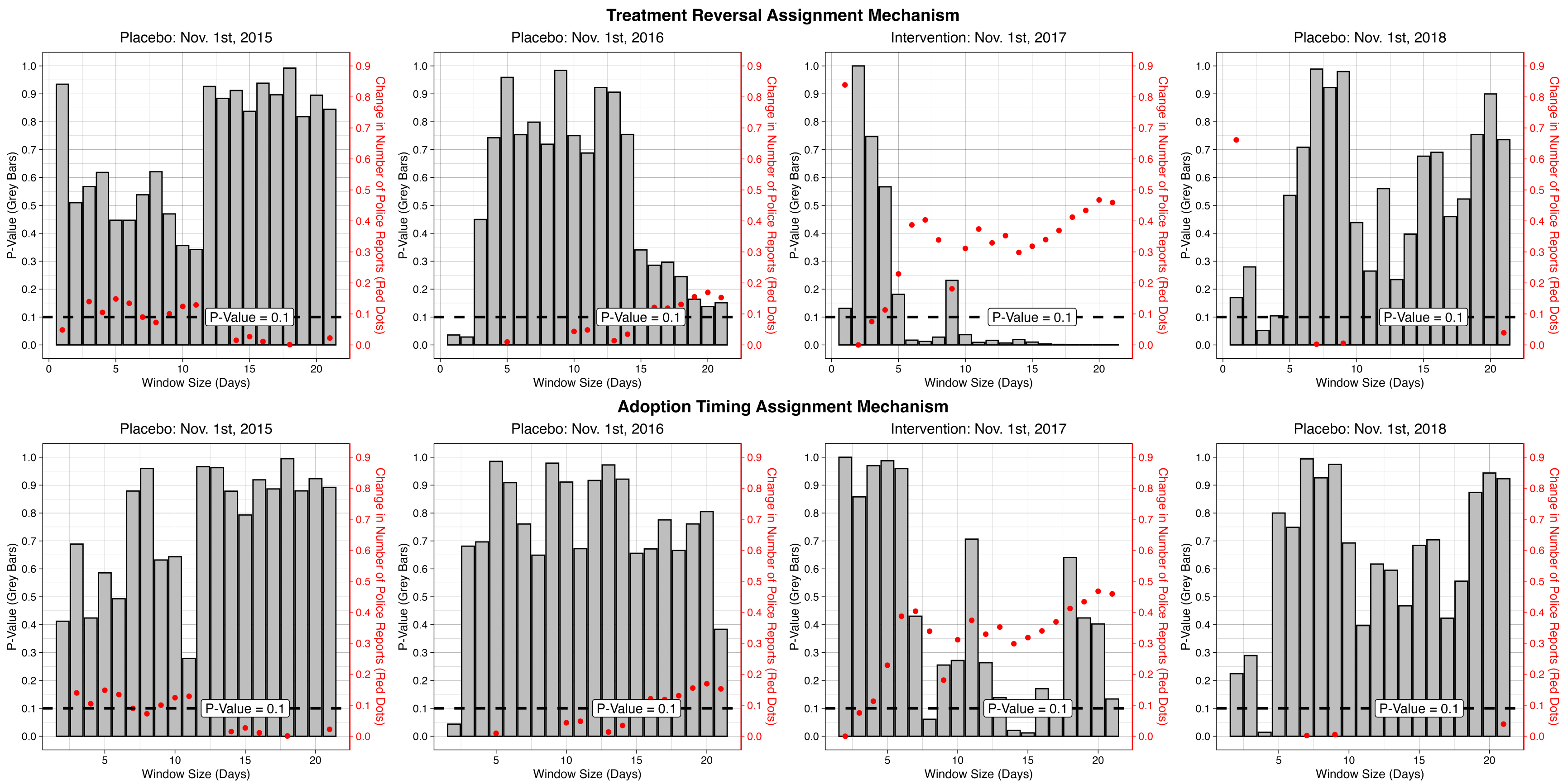


Figure SA-6: Effects of CCP Reform for Different Windows Around *Placebo* Adoption Times (*Day*)

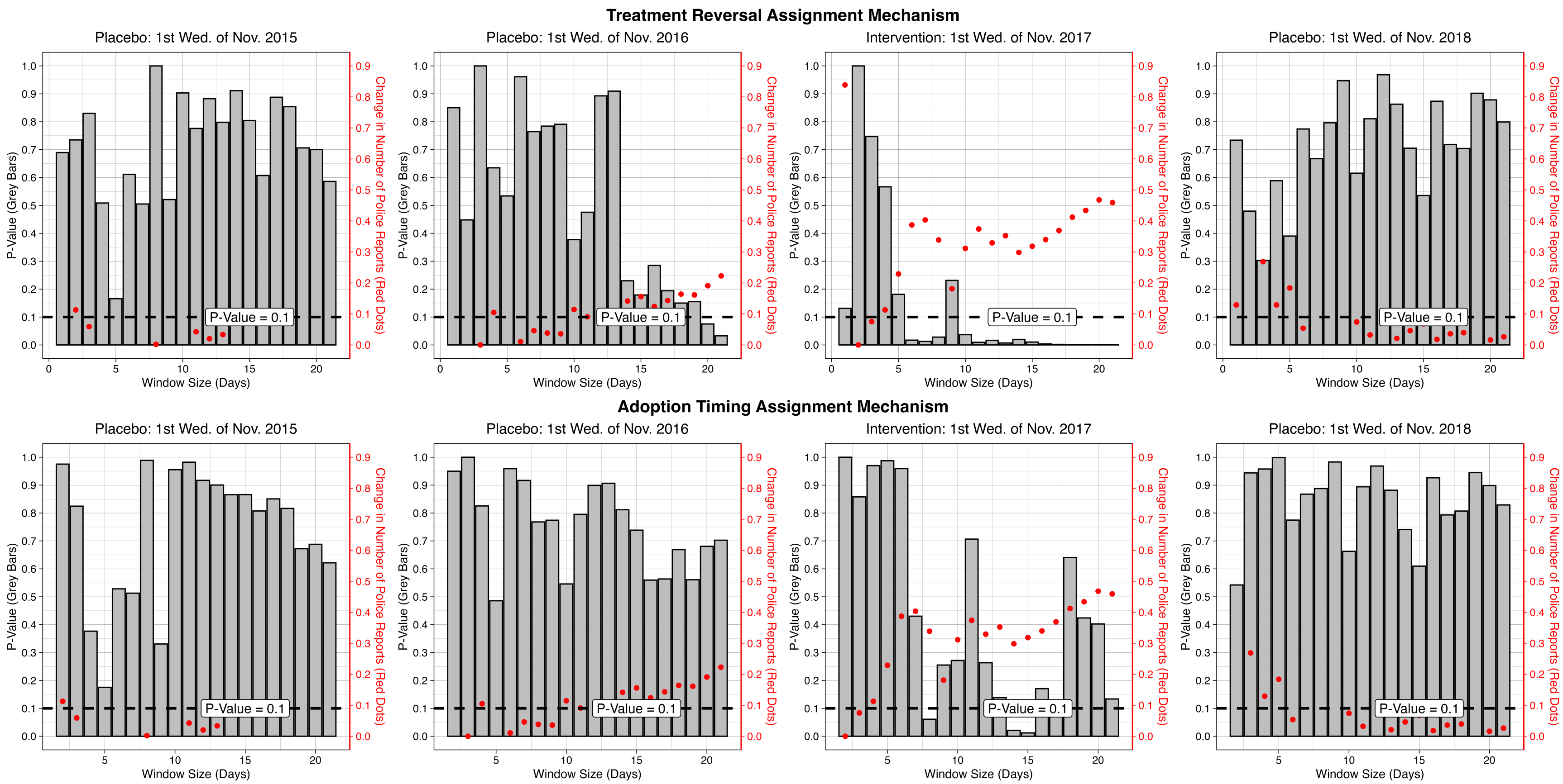


Figure SA-7: Effects of CCP Reform When Restricting *Reporting Delay*

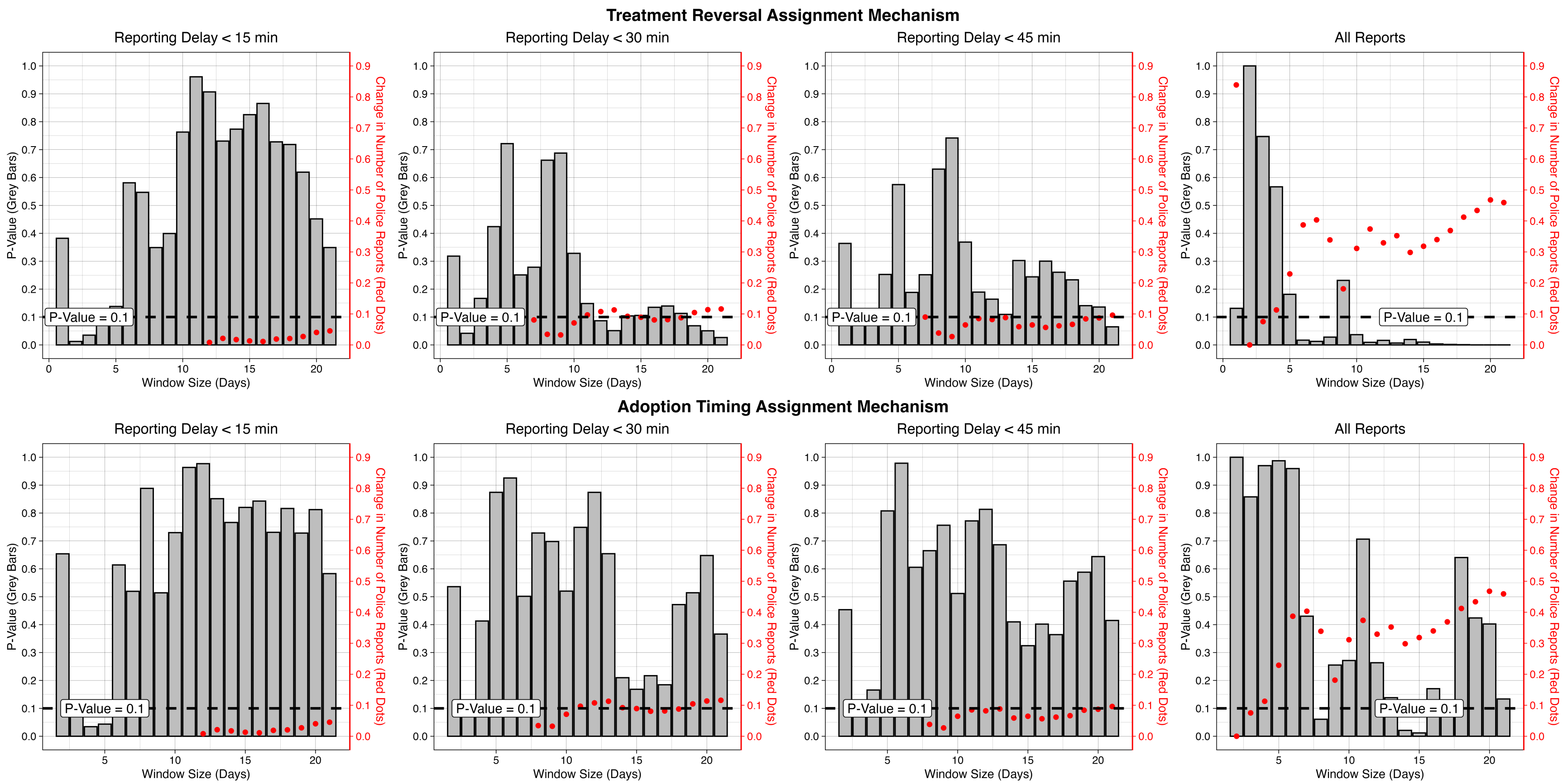


Figure SA-8: Effects of CCP Reform for Different Windows Around *Placebo* Adoption Times, Additional Years (*Date*)