ONLINE APPENDIX MATERIALS FOR:

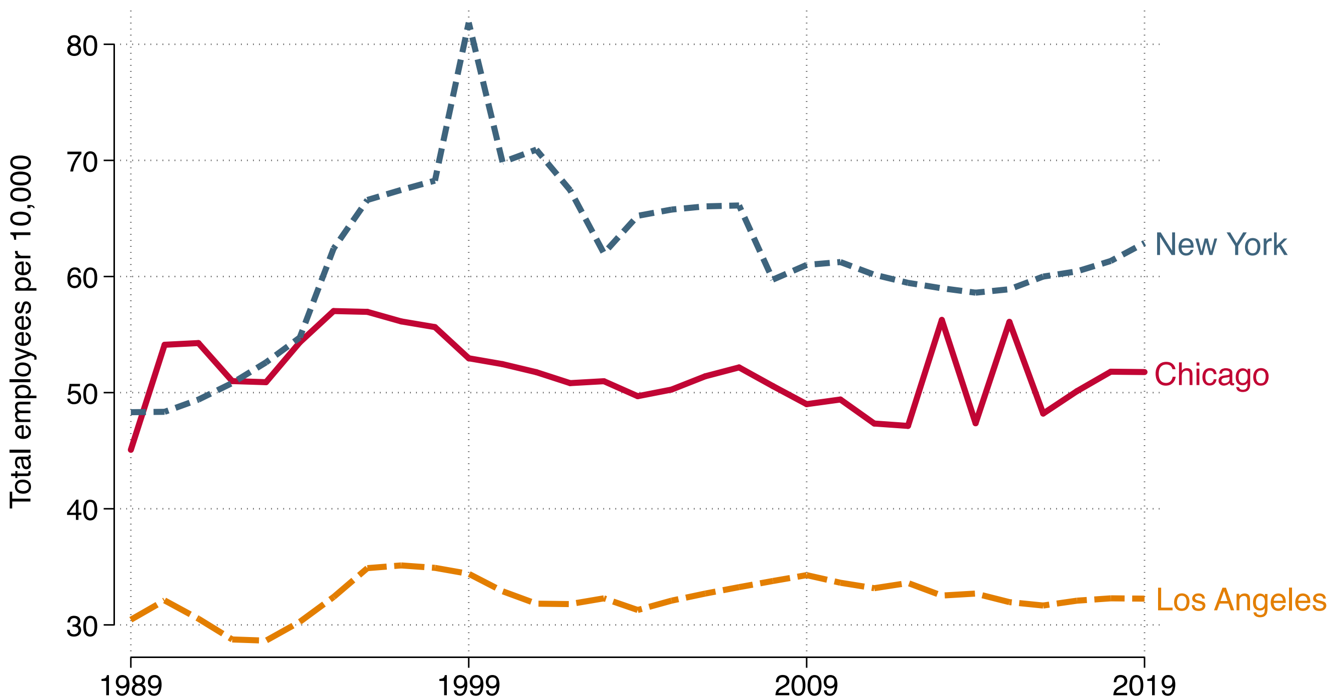
POLICING AND MANAGEMENT

March 1, 2022

Max Kapustin
Terrence D. Neumann
Jens Ludwig

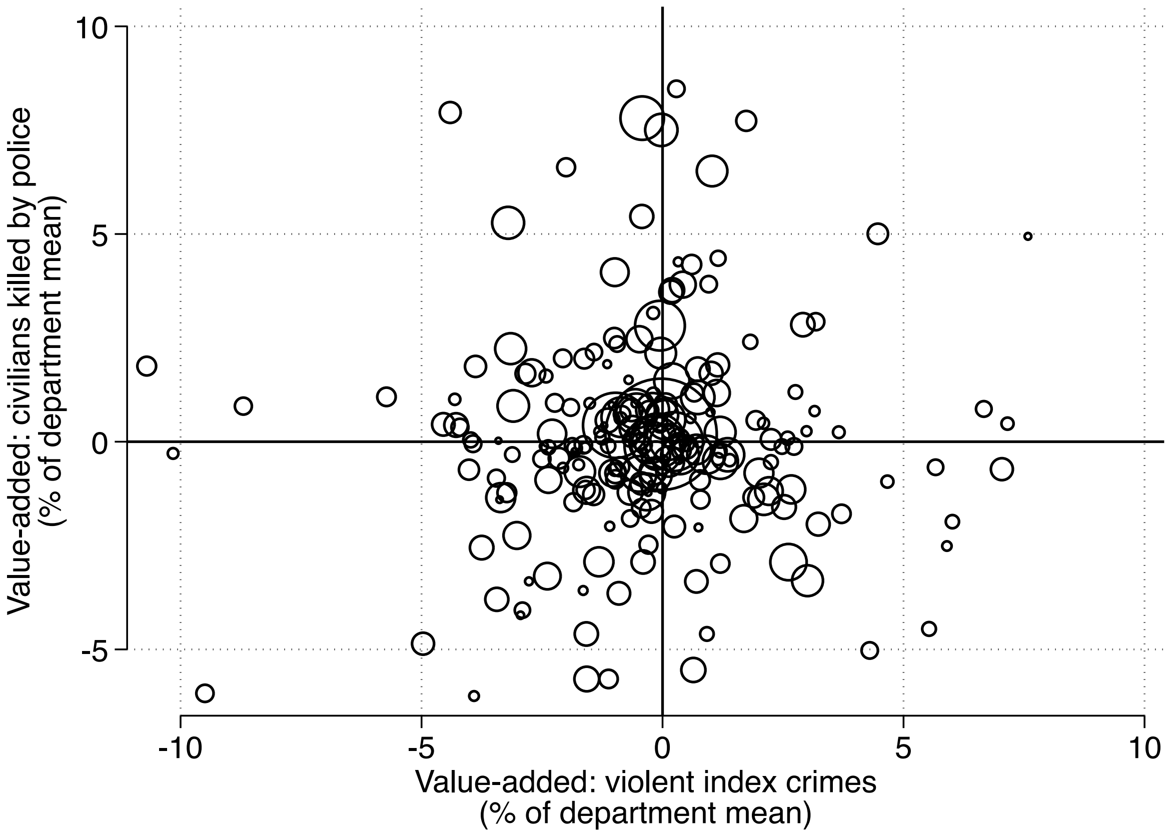
Appendix

Figure 1: Police employees (sworn & civilian) in New York City, Los Angeles, and Chicago, 1989–2019



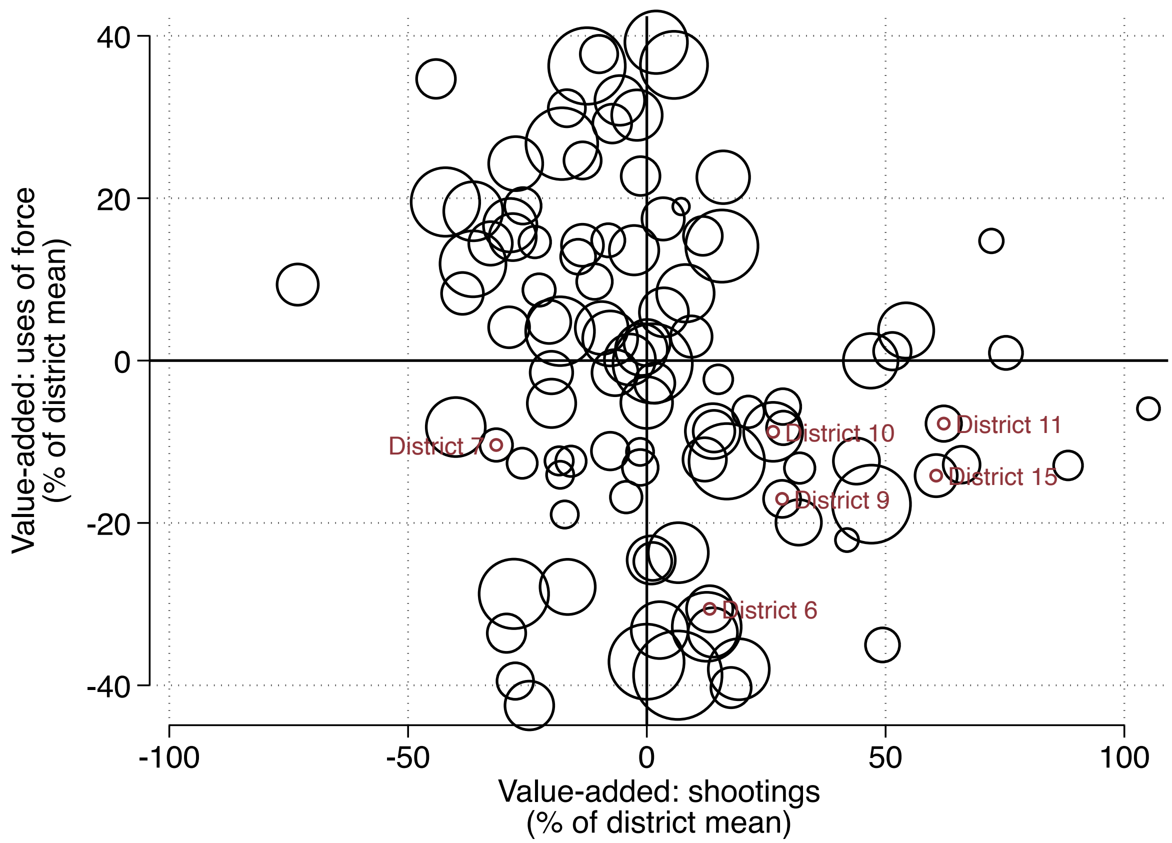
Note: Data from UCR LEOKA and NYPD Office of Management Analysis and Planning (OMAP). NYPD sworn staffing levels from 1990-2009 are based on OMAP data made available by Franklin Zimring (`https://global.oup.com/us/companion.websites/9780199844425/`). For discussion of errors in NYPD's sworn staffing levels in UCR data, see Chalfin and McCrary (2018).

Figure 2: Within-city, across-chief variation: value-added



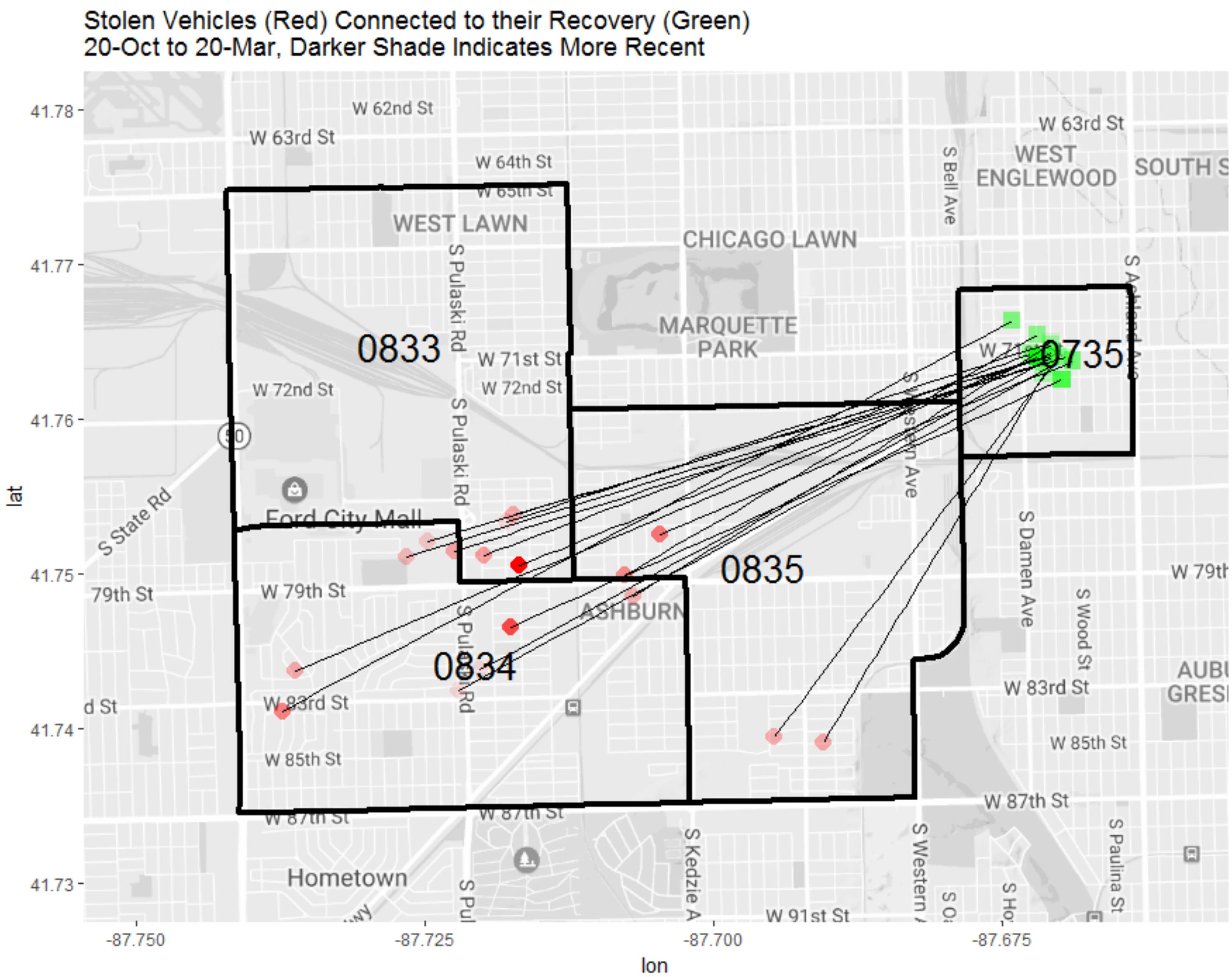
Note: Data from UCR and Fatal Encounters. Departments serving the 50 largest jurisdictions based on median population in 2010-2019. Each point is a pair of value-added effect estimates obtained using the shrinkage estimator in Easterly and Pennings (2020). The sizes of the points reflect weighting for both jurisdiction population and tenure length. Value-added effects are not estimated for chiefs with tenures of less than 6 months or who are serving in an interim capacity. Points where either value-added effect estimate is above (below) the 99th (1st) percentile are not plotted. On the x-axis is the chief's estimated effect on violent index crime rates as a percent of their department mean violent index crime rate. On the y-axis is the chief's estimated effect on civilian-police death rates as a percent of their department mean civilian-police death rate.

Figure 3: Within-district, across-commander variation: tenure fixed effects



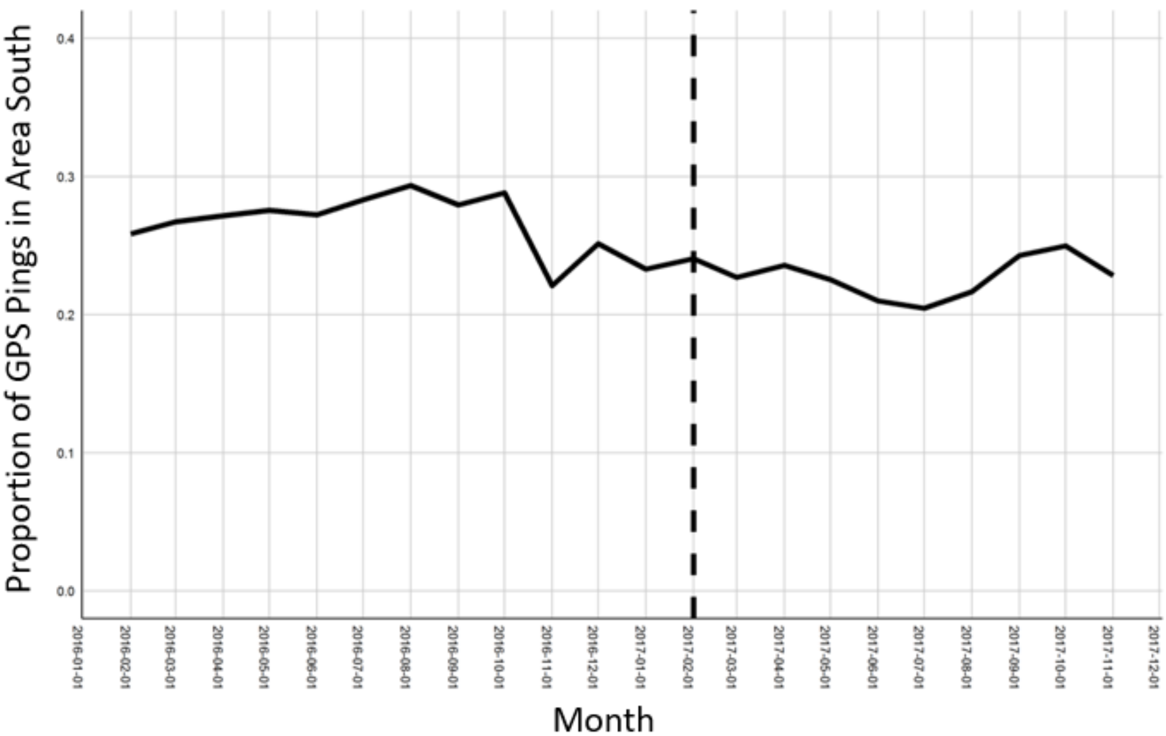
Note: Data from the Chicago Police Department. Each point is a pair of value-added effect estimates obtained using the shrinkage estimator in Easterly and Pennings (2020). The sizes of the points reflect weighting for both district population and tenure length. Fixed effects are not estimated for commanders with tenures of less than 6 months or who are serving in an interim capacity. Points where either value-added effect estimate is above (below) the 99th (1st) percentile are not plotted. On the x-axis is the commander's estimated effect on shooting rates as a percent of their district mean shooting rate. On the y-axis is the commander's estimated effect on use of force rates as a percent of their district mean use of force rate. The six labeled points represent the tenures of the commanders of the six Tier 1 districts during most or all of 2017.

Figure 4: Sample SDSC analysis involving motor vehicle thefts from the 7th district



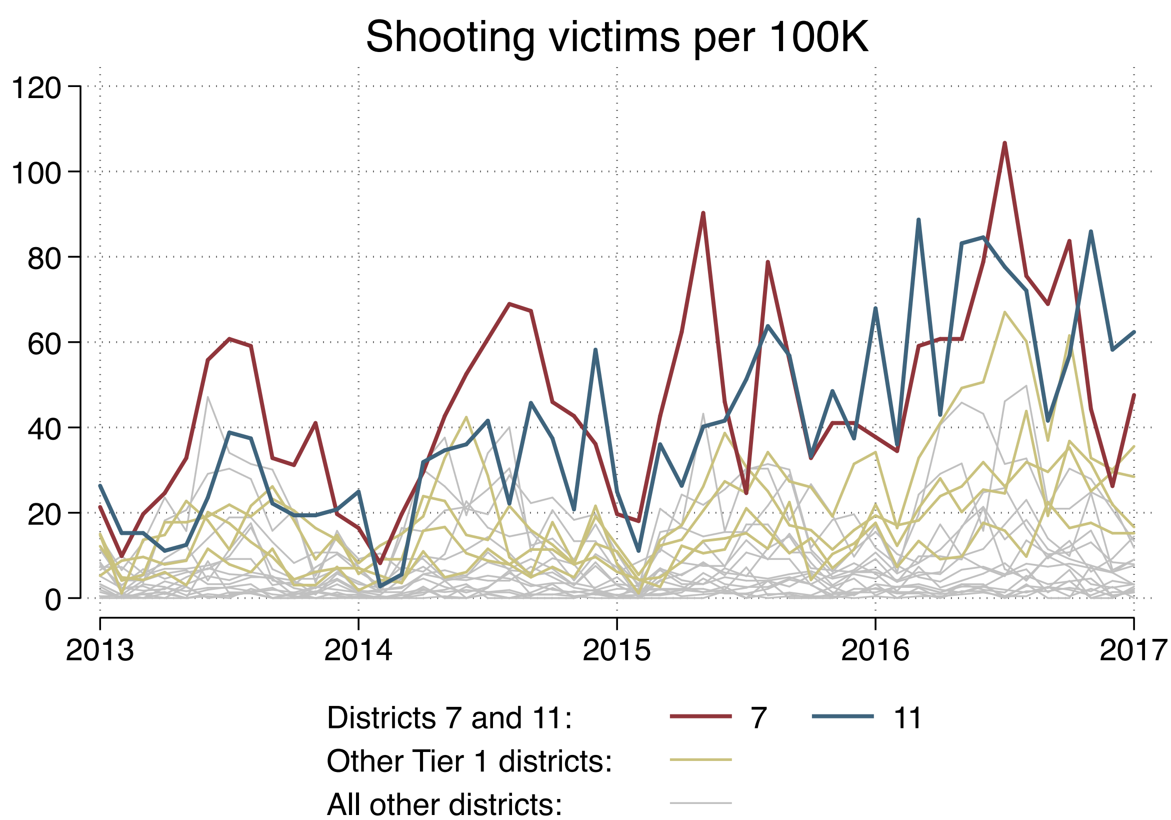
Note: Red dots represent motor vehicle theft locations. Green dots represent motor vehicle theft recovery locations. See section 3 for details.

Figure 5: Officer time in the 7th district



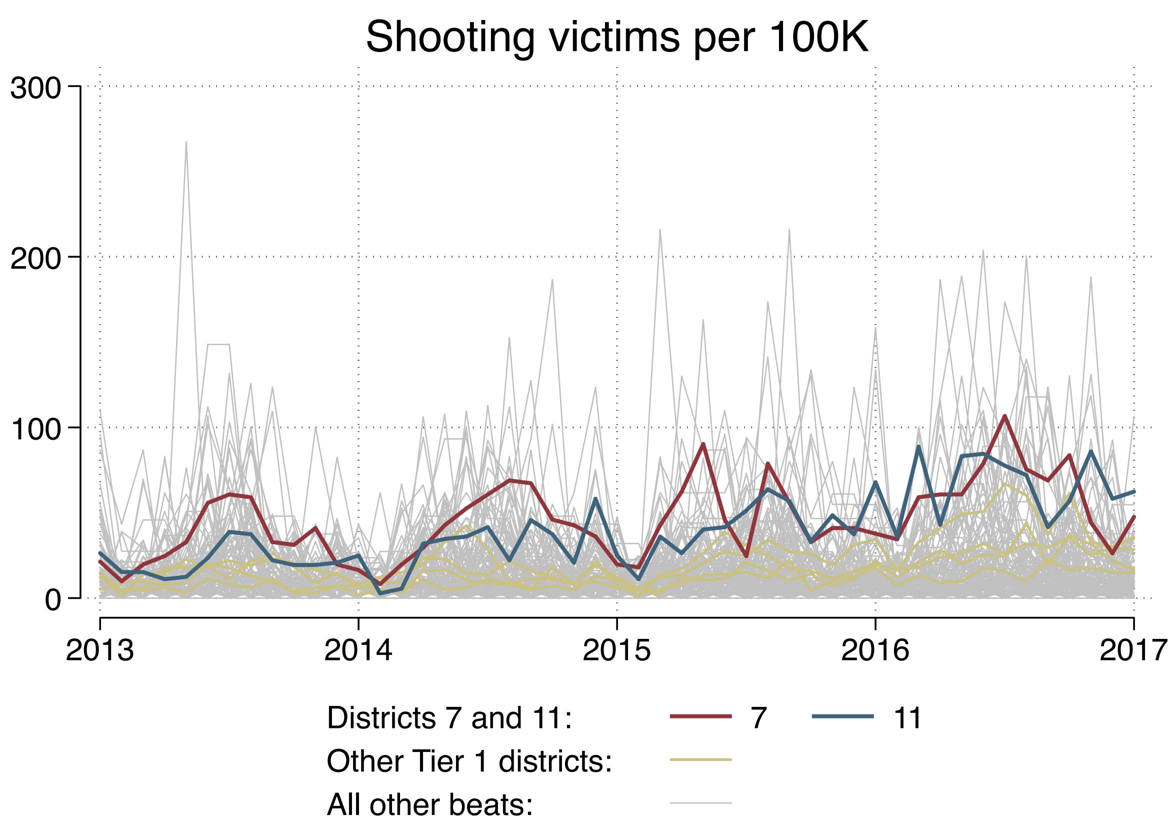
Note: Data on officer vehicle GPS pings within the 7th district from the Chicago Police Department.

Figure 6: Shooting victimization rate, 2013–2017: Tier 1 districts vs. other districts



Note: Data from the Chicago Police Department.

Figure 7: Shooting victimization rate, 2013–2017: Tier 1 districts vs. other beats



Note: Data from the Chicago Police Department.

Figure 8: Distribution of effect estimates across model specifications

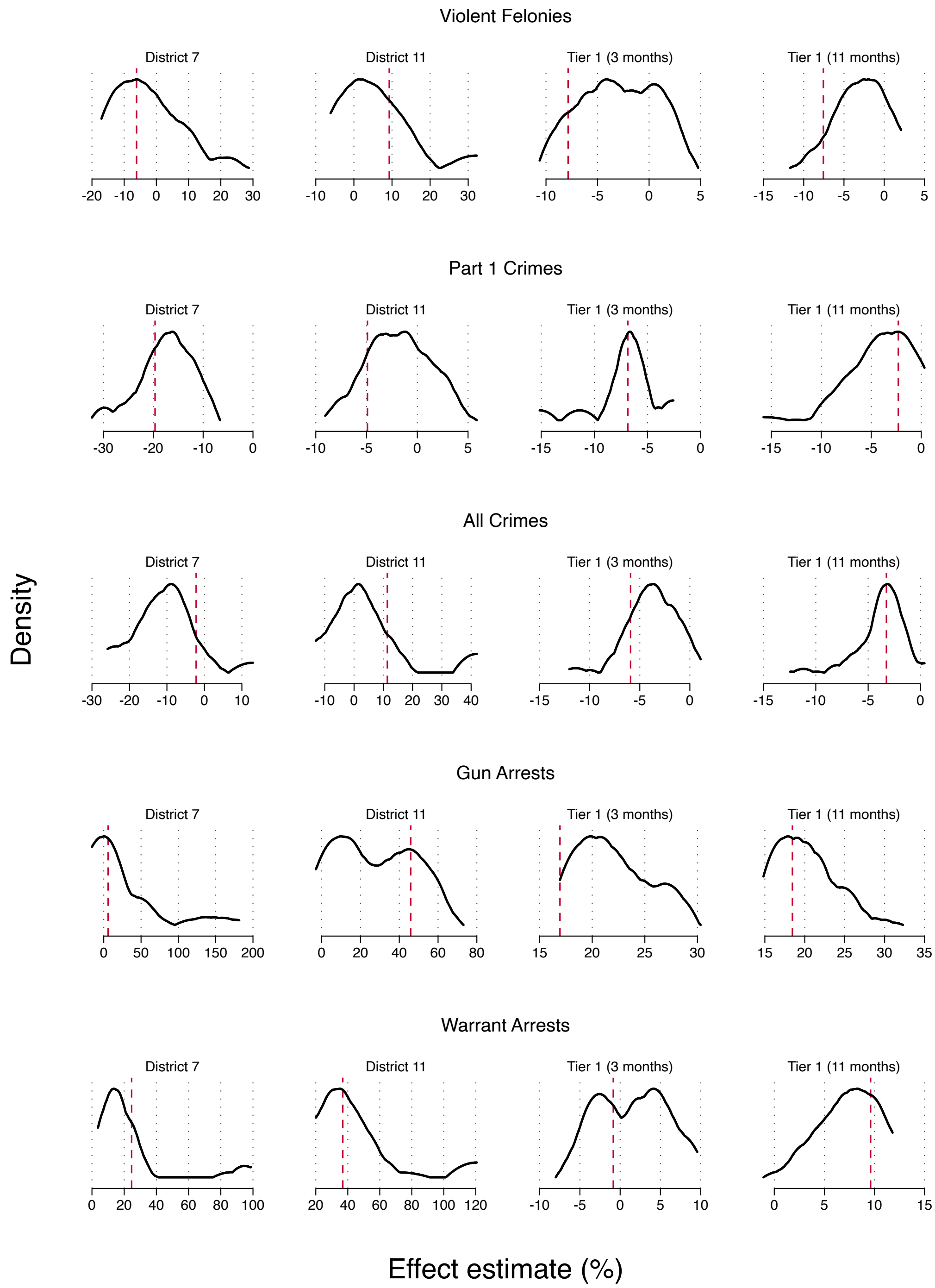
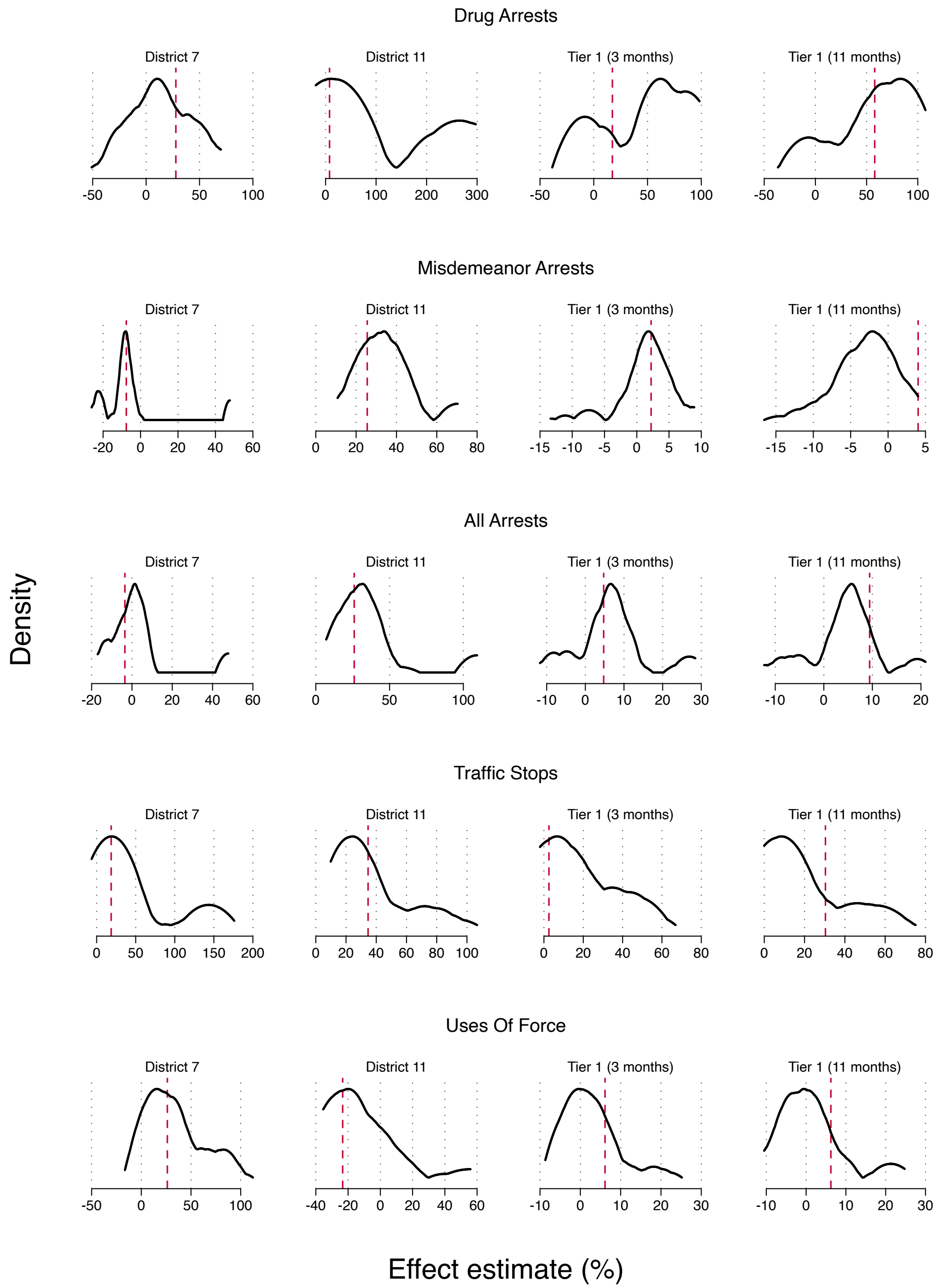
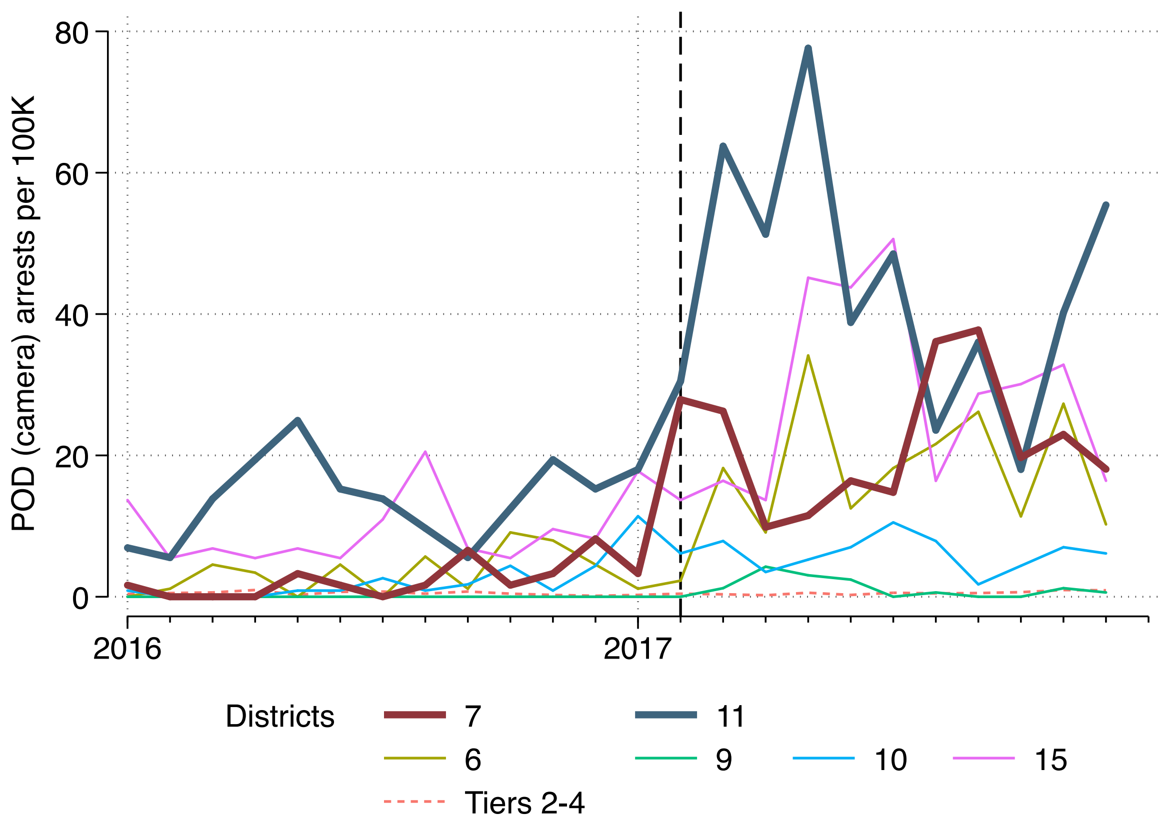


Figure 8: Distribution of effect estimates across model specifications (cont.)



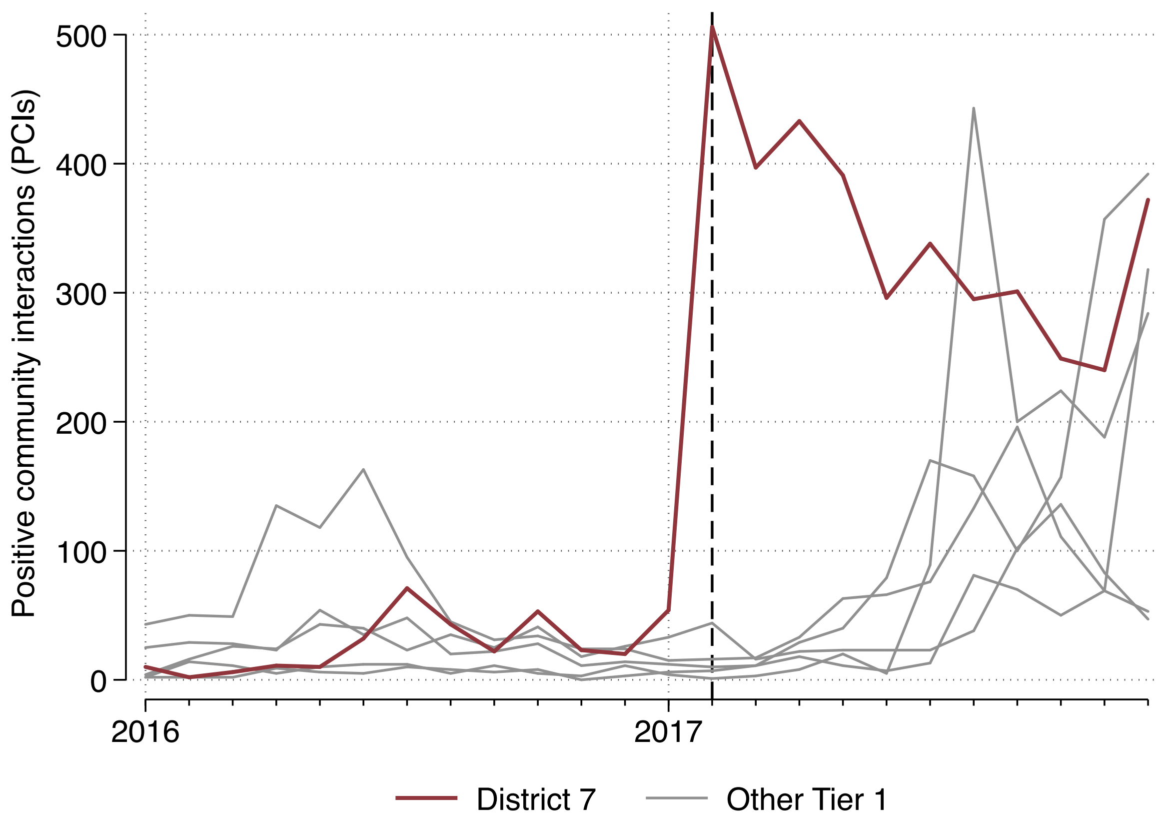
Note: Data from the Chicago Police Department. Each panel reports distributions of estimated effects on the indicated outcome from different model specifications. Vertical lines correspond to the model specification that maximizes out-of-sample prediction accuracy in the backdating exercise.

Figure 9: Police camera-initiated arrests



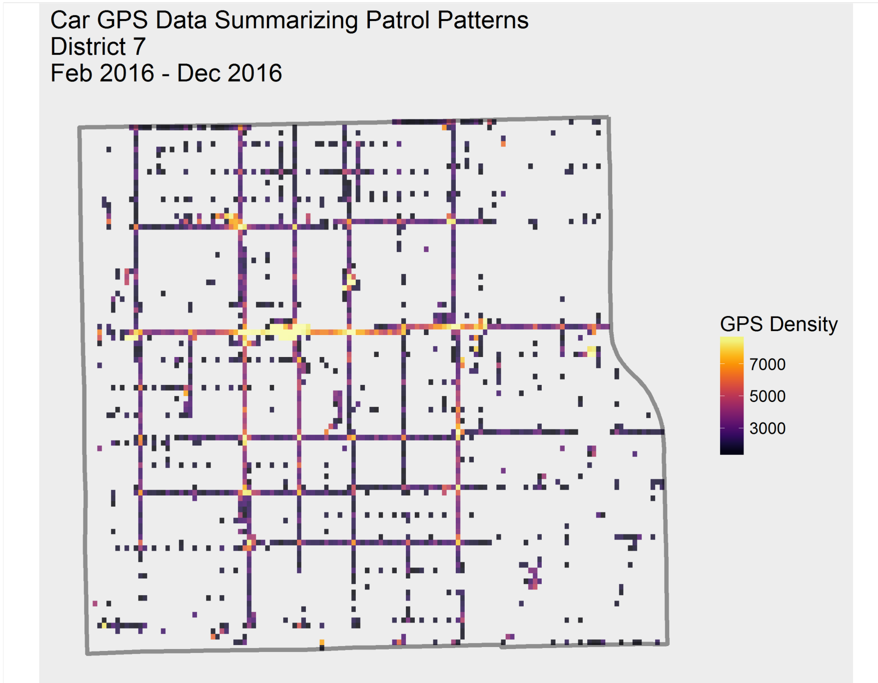
Note: Data from the Chicago Police Department. Rates of arrests initiated by officers monitoring Police Observation Device (POD) cameras, for the 7^{th} and 11^{th} districts, the four other Tier 1 districts, and all other police districts (Tiers 2-4). Dashed vertical line indicates the introduction of SDSCs in the 7^{th} and 11^{th} districts in February 2017. SDSCs were introduced in the four other Tier 1 districts in March 2017, and in other districts in 2018 and later.

Figure 10: Positive community interactions (PCIs)

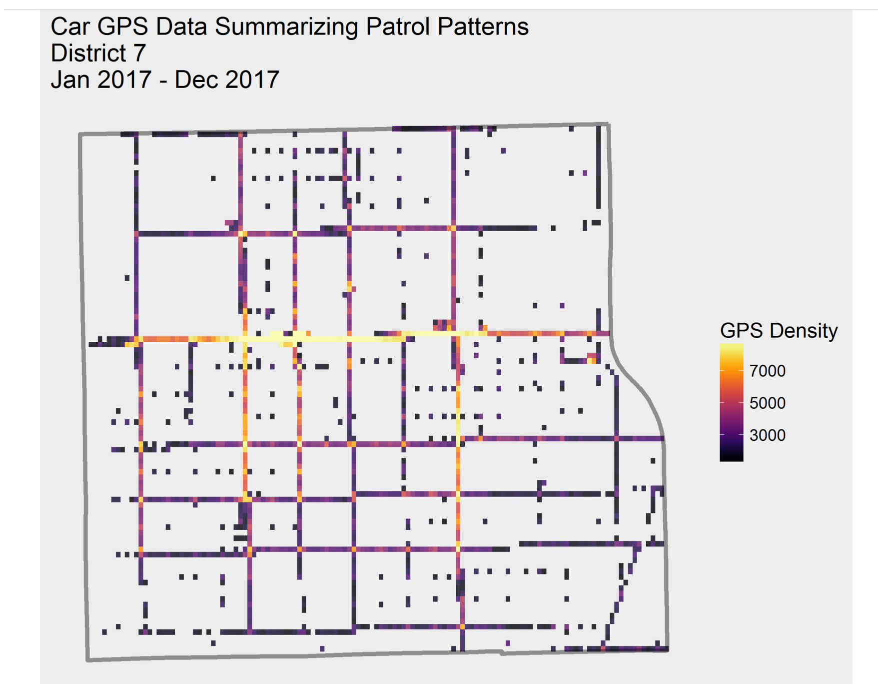


Note: Data from the Chicago Police Department. Monthly volume of positive community interactions (PCIs) by officers in Tier 1 districts. Dashed vertical line indicates the introduction of the SDSC in the 7^{th} districts in February 2017.

Figure 11: Patrol patterns in the 7th district



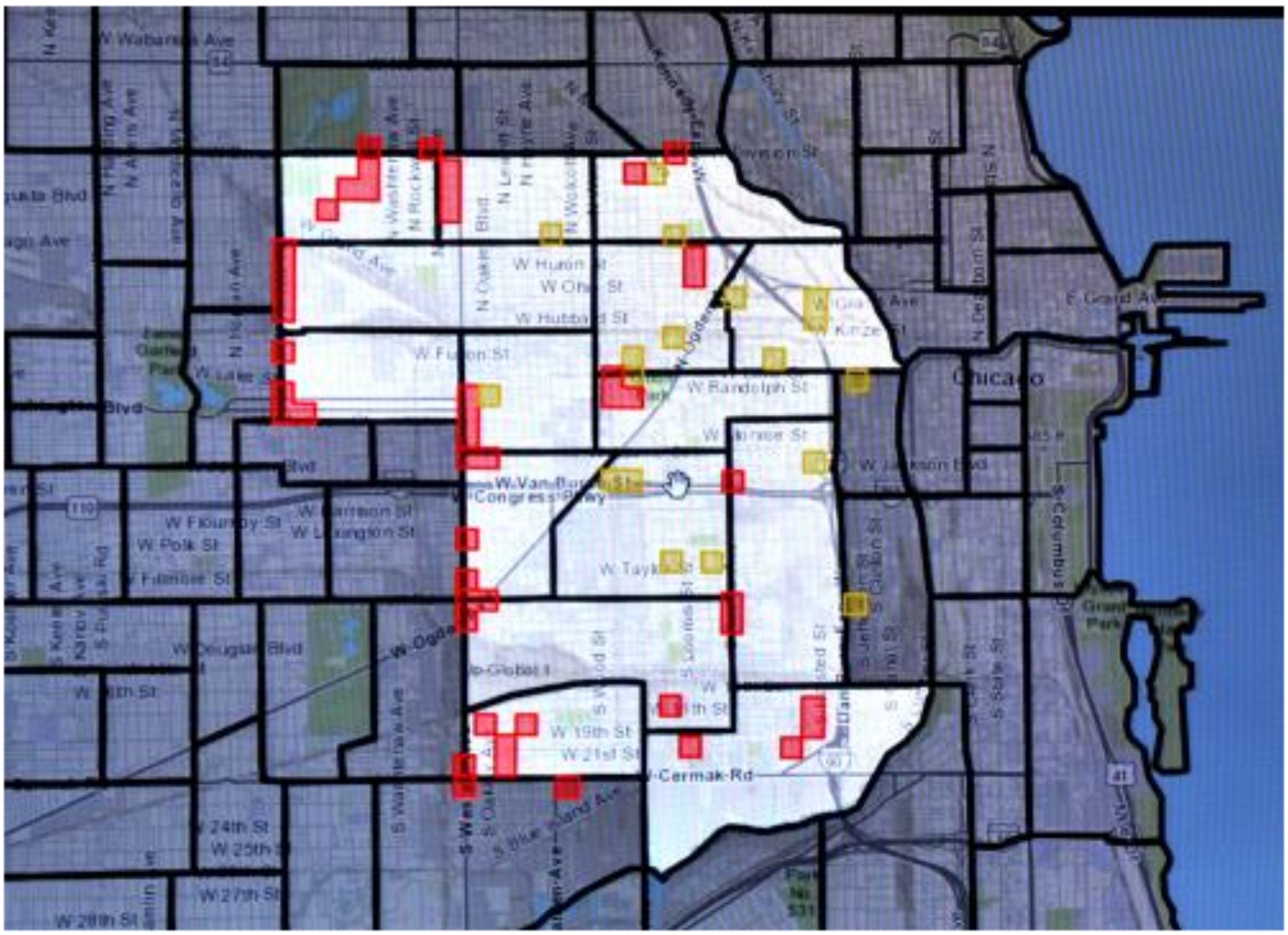
(a) 2016



(b) 2017

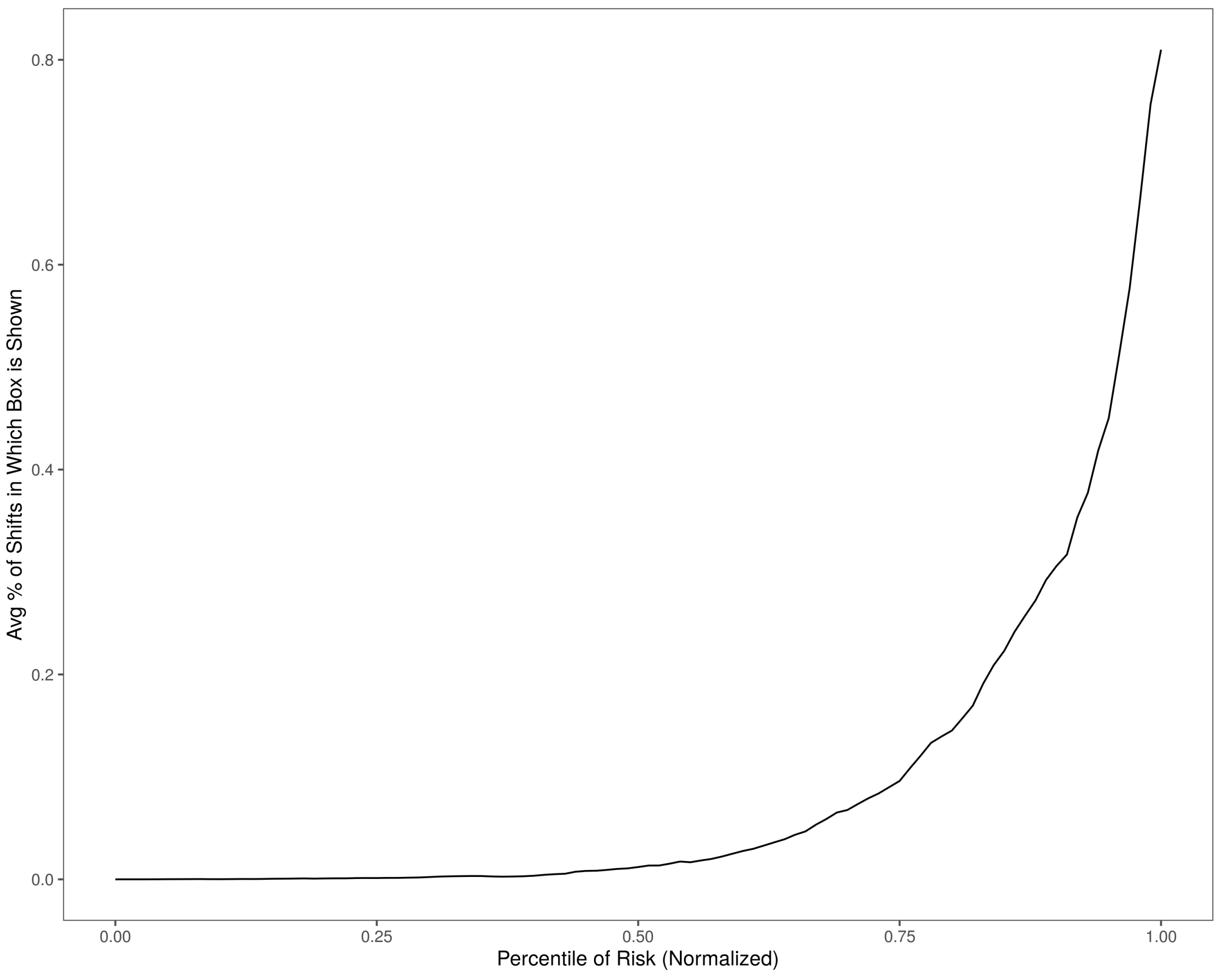
Note: Data from officer vehicle GPS pings within the 7th district. Figure shows the top 15% of cells by frequency of GPS pings in February 2016 - December 2016 and January 2017 - December 2017, respectively. Cells shaded lighter saw a greater frequency of GPS pings. In both periods, the area near the 7th district station, located at W 63rd St and S Loomis Blvd, has among the greatest frequency of GPS pings.

Figure 12: Example: HunchLab in the 12$^{\text{th}}$ district



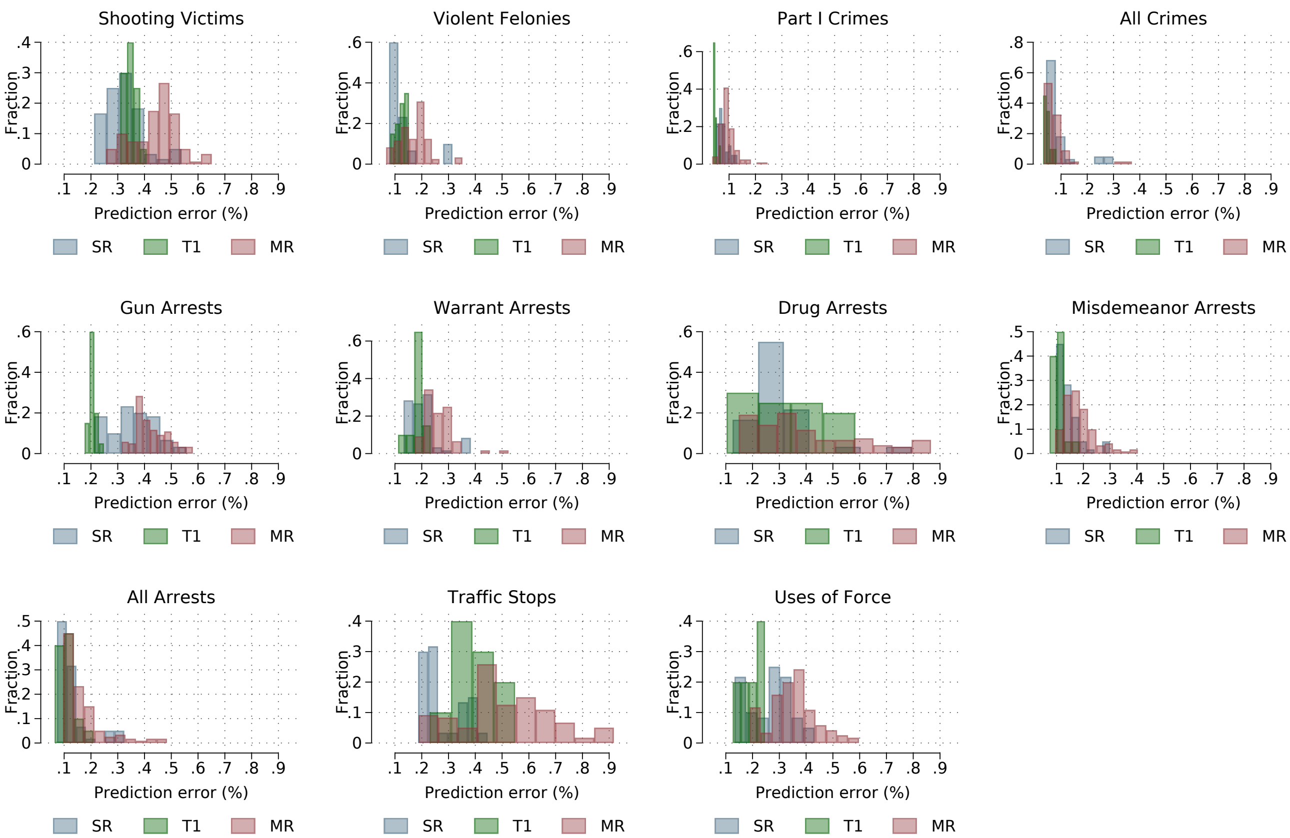
Note: The broader highlighted region is the 12$^{\text{th}}$ district. The dark borders within the district outline police beats, each of which have four boxes shown per watch, in red. See section 5 for details.

Figure 13: Estimated probability of a HunchLab box being shown vs. its relative weighted risk



Note: Data from the Chicago Police Department. Estimated probability of a box being shown by HunchLab using a moving average across lotteries (y-axis) and a box's relative weighted risk normalized within its beat (x-axis). See equation 8.

Figure 14: Prediction errors from backdating exercise, by analysis type



Note: Histograms of prediction errors (%) from the backdating exercise across model specifications. SR = short-run analysis of the 7th and 11th districts in February 2017. T1 = analysis of Tier 1 districts as a whole through December 2017. MR = medium-run analysis of individual Tier 1 districts through December 2017. Model specifications vary by estimator (DD, SCM, ASCM, EN, EN-M), donor pool, and donor unit type (district, beat). For each specification, treated unit, and outcome, we estimate models using data from the first $T_0^{\text{placebo}} \geq \underline{T_0}^{\text{placebo}}$ pre-treatment months, assess their fit (RMSPE) in the last $T_0 - T_0^{\text{placebo}}$ pre-treatment months, and scale this by the average outcome in the last $T_0 - T_0^{\text{placebo}}$ pre-treatment months to obtain a prediction error (%). The distributions of these prediction errors, separately by outcome and analysis type, are shown. For additional details about the backdating exercise, see Appendix B.1.

Table 1: Balance table: HunchLab full experimental sample

	6th district			7th district			9th district		
Predicted risk	Shown	Not Shown	*p*-value	Shown	Not Shown	*p*-value	Shown	Not Shown	*p*-value
Weighted	0.0548	0.0226	0.0000	0.0592	0.0322	0.0000	0.0396	0.0122	0.0000
Homicide	0.0001	0.0001	0.0000	0.0001	0.0001	0.0000	0.0001	0.0000	0.0000
Shooting	0.0006	0.0002	0.0000	0.0007	0.0003	0.0000	0.0005	0.0001	0.0000
Robbery	0.0013	0.0005	0.0000	0.0010	0.0005	0.0000	0.0006	0.0002	0.0000
Assault	0.0004	0.0002	0.0000	0.0004	0.0003	0.0000	0.0003	0.0001	0.0000
N	22740	263082		28441	207611		28440	450300	

	10th district			11th district			15th district		
Predicted risk	Shown	Not Shown	*p*-value	Shown	Not Shown	*p*-value	Shown	Not Shown	*p*-value
Weighted	0.0690	0.0211	0.0000	0.1030	0.0402	0.0000	0.1106	0.0427	0.0000
Homicide	0.0001	0.0000	0.0000	0.0002	0.0001	0.0000	0.0002	0.0001	0.0000
Shooting	0.0010	0.0003	0.0000	0.0015	0.0005	0.0000	0.0017	0.0005	0.0000
Robbery	0.0009	0.0003	0.0000	0.0018	0.0008	0.0000	0.0016	0.0008	0.0000
Assault	0.0004	0.0002	0.0000	0.0005	0.0002	0.0000	0.0005	0.0002	0.0000
N	22462	251984		26584	182450		16114	109970	

Note: Data from the Chicago Police Department. 36,024 lotteries taking place between May 15, 2017 and October 31, 2017.

Table 2: Balance table: HunchLab analysis sample

	6th district			7th district			9th district		
Predicted risk	Shown	Not Shown	*p*-value	Shown	Not Shown	*p*-value	Shown	Not Shown	*p*-value
Weighted	0.0443	0.0445	0.68018	0.0517	0.0517	0.94867	0.0295	0.0294	0.76347
Homicide	0.0001	0.0001	0.14488	0.0001	0.0001	0.12584	0.0001	0.0001	0.47011
Shooting	0.0005	0.0005	0.67979	0.0006	0.0006	0.36180	0.0003	0.0003	0.34241
Robbery	0.0010	0.0010	0.39054	0.0009	0.0009	0.16890	0.0005	0.0005	0.04537
Assault	0.0004	0.0004	0.16761	0.0004	0.0004	0.18486	0.0003	0.0003	0.25619
N	5887	5887		5790	5790		7852	7852	

	10th district			11th district			15th district		
Predicted risk	Shown	Not Shown	*p*-value	Shown	Not Shown	*p*-value	Shown	Not Shown	*p*-value
Weighted	0.0501	0.0498	0.56019	0.0794	0.0794	0.93970	0.0853	0.0847	0.59474
Homicide	0.0001	0.0001	0.41729	0.0002	0.0002	0.22369	0.0002	0.0002	0.67289
Shooting	0.0007	0.0007	0.13652	0.0011	0.0011	0.36985	0.0012	0.0012	0.23989
Robbery	0.0007	0.0007	0.48189	0.0014	0.0014	0.41580	0.0015	0.0015	0.42169
Assault	0.0003	0.0003	0.09744	0.0004	0.0004	0.10341	0.0004	0.0004	0.19794
N	6262	6262		5827	5827		3266	3266	

Note: Data from the Chicago Police Department.

Table 3: Estimated effects of HunchLab on officer time in a box, and of officer time on ShotSpotter alerts, by district (no buffer)

	Officer minutes			ShotSpotter alerts		
District	Control mean (analysis sample)	Effect of showing box	First-stage F-statistic	Control mean (full sample)	Control mean (analysis sample)	Effect of minute of officer time
	(1)	(2)	(3)	(4)	(5)	(6)
Pooled	10.05	0.2004**	6.16	0.0017	0.0032	-0.0009
		(0.0808)				(0.0025)
						[-0.0131,0.0113]
6	9.20	0.2675	1.03	0.0001	0.0002	0.0007
		(0.2633)				(0.0009)
7	11.48	0.4252**	6.51	0.0031	0.0040	-0.0036
		(0.1667)				(0.0029)
						[-0.0173,0.0101]
9	8.00	0.2866**	4.47	0.0009	0.0023	-0.0039
		(0.1355)				(0.0030)
						[-0.0314,0.0235]
10	8.93	0.1473	0.70	0.0002	0.0004	-0.0005
		(0.1761)				(0.0032)
11	10.28	0.1327	0.61	0.0045	0.0086	0.0120
		(0.1697)				(0.0235)
15	14.50	-0.2591	0.53	0.0027	0.0025	-0.0034
		(0.3574)				(0.0079)

Note: Data from the Chicago Police Department. ShotSpotter alerts defined as only those occurring within a HunchLab box. First and second stage estimates from equation 9. Heteroskedasticity-robust standard errors clustered at the box level reported in parentheses. tF-adjusted 95% confidence intervals calculated using the method described in Lee et al. (2021) reported in brackets where first-stage F-statistic exceeds 4. See section 5 for details. * $p < 0.1$ ** $p < 0.05$ *** $p < 0.01$.

Table 4(a): Backdating exercise results: 7th and 11th districts, short-run

District	Outcome	Prediction error (%)		Best-performing specification		
		Maximum	Minimum	Estimator	Donor pool	Donor type
7	Shooting Victims	53.9	27.5	EN-M	Tier 1-4	Districts
11	Shooting Victims	40.8	21.2	EN	Tier 1-2	Beats
7	Violent Felonies	31.4	8.1	ASCM	Tier 1-2	Districts
11	Violent Felonies	31.4	7.8	EN-M	Tier 1	Districts
7	Part 1 Crimes	13.2	6.4	EN-M	Tier 1-4	Districts
11	Part 1 Crimes	8.0	5.1	SCM	Tier 1-4	Beats
7	All Crimes	25.0	6.5	SCM	Tier 1-2	Beats
11	All Crimes	30.0	4.3	ASCM	Tier 1-2	Beats
7	Gun Arrests	43.1	21.3	ASCM	Tier 1-4	Districts
11	Gun Arrests	55.7	35.1	ASCM	Tier 1-2	Districts
7	Warrant Arrests	35.5	13.2	EN-M	Tier 1-2	Districts
11	Warrant Arrests	38.3	16.6	EN-M	Tier 1	Districts
7	Drug Arrests	79.5	21.0	EN-M	Tier 1-4	Districts
11	Drug Arrests	37.8	12.5	EN	Tier 1-2	Districts
7	Misdemeanor Arrests	30.0	11.1	EN	Tier 1-4	Beats
11	Misdemeanor Arrests	18.4	9.8	EN-M	Tier 1-4	Districts
7	All Arrests	25.4	8.1	EN	Tier 1-2	Beats
11	All Arrests	32.4	7.3	EN	Tier 1	Beats
7	Traffic Stops	40.6	18.7	EN-M	Tier 1-2	Beats
11	Traffic Stops	44.6	21.0	EN	Tier 1-4	Districts
7	Uses Of Force	37.6	26.2	EN-M	Tier 1-4	Districts
11	Uses Of Force	43.1	13.3	EN	Tier 1-2	Districts

Note: Data from the Chicago Police Department.

Table 4(b): Backdating exercise results: Tier 1

		Prediction error (%)		Best-performing specification		
Interval	Outcome	Maximum	Minimum	Estimator	Donor pool	Donor type
1 month	Shooting Victims	29.2	20.5	EN	Tier 2	Beats
2 months	Shooting Victims	33.1	23.1	EN	Tier 2	Beats
3 months	Shooting Victims	34.9	25.6	EN	Tier 2	Beats
6 months	Shooting Victims	37.2	28.2	EN	Tier 2	Beats
11 months	Shooting Victims	40.8	31.0	SCM	Tier 2	Beats
1 month	Violent Felonies	11.3	6.0	EN	Tier 2-4	Beats
2 months	Violent Felonies	12.5	6.8	EN	Tier 2-4	Districts
3 months	Violent Felonies	13.1	7.0	EN	Tier 2-4	Districts
6 months	Violent Felonies	13.8	7.5	EN	Tier 2-4	Districts
11 months	Violent Felonies	15.2	8.2	EN-M	Tier 2-4	Districts
1 month	Part 1 Crimes	5.1	2.8	DD	Tier 2	Districts
2 months	Part 1 Crimes	5.5	3.2	DD	Tier 2	Districts
3 months	Part 1 Crimes	5.7	3.3	DD	Tier 2	Districts
6 months	Part 1 Crimes	6.3	3.6	DD	Tier 2	Districts
11 months	Part 1 Crimes	6.9	3.9	EN	Tier 2	Districts
1 month	All Crimes	6.7	2.9	EN	Tier 2	Districts
2 months	All Crimes	6.7	3.3	DD	Tier 2	Beats
3 months	All Crimes	6.7	3.2	DD	Tier 2	Beats
6 months	All Crimes	7.0	3.2	EN	Tier 2	Districts
11 months	All Crimes	8.2	3.2	EN	Tier 2	Districts

Table 4(b): Backdating exercise results: Tier 1 (cont.)

		Prediction error (%)		Best-performing specification		
Interval	Outcome	Maximum	Minimum	Estimator	Donor pool	Donor type
1 month	Gun Arrests	20.8	13.1	DD	Tier 2	Districts
2 months	Gun Arrests	21.8	15.5	DD	Tier 2	Districts
3 months	Gun Arrests	23.2	16.5	DD	Tier 2	Districts
6 months	Gun Arrests	25.9	17.9	DD	Tier 2	Districts
11 months	Gun Arrests	24.8	17.6	ASCM	Tier 2-4	Beats
1 month	Warrant Arrests	17.2	10.7	EN	Tier 2-4	Districts
2 months	Warrant Arrests	18.0	11.6	EN	Tier 2-4	Districts
3 months	Warrant Arrests	18.3	11.8	EN	Tier 2-4	Districts
6 months	Warrant Arrests	18.7	13.2	EN	Tier 2-4	Districts
11 months	Warrant Arrests	23.3	11.2	EN-M	Tier 2-4	Districts
1 month	Drug Arrests	57.3	13.1	EN	Tier 2-4	Districts
2 months	Drug Arrests	55.8	14.0	EN	Tier 2-4	Districts
3 months	Drug Arrests	54.8	15.3	EN	Tier 2-4	Districts
6 months	Drug Arrests	54.3	17.4	ASCM	Tier 2-4	Districts
11 months	Drug Arrests	58.3	10.3	EN-M	Tier 2-4	Districts
1 month	Misdemeanor Arrests	14.3	5.6	SCM	Tier 2-4	Beats
2 months	Misdemeanor Arrests	14.5	6.5	SCM	Tier 2-4	Beats
3 months	Misdemeanor Arrests	14.7	7.0	SCM	Tier 2-4	Beats
6 months	Misdemeanor Arrests	15.7	7.5	EN	Tier 2-4	Districts
11 months	Misdemeanor Arrests	18.5	7.4	EN-M	Tier 2	Districts
1 month	All Arrests	15.9	5.1	EN	Tier 2	Districts
2 months	All Arrests	15.9	5.9	EN	Tier 2-4	Districts
3 months	All Arrests	16.0	6.2	EN	Tier 2-4	Districts
6 months	All Arrests	17.3	6.2	EN	Tier 2	Districts
11 months	All Arrests	20.9	6.4	EN-M	Tier 2	Districts

Table 4(b): Backdating exercise results: Tier 1 (cont.)

Interval	Outcome	Prediction error (%)		Best-performing specification		
		Maximum	Minimum	Estimator	Donor pool	Donor type
1 month	Traffic Stops	29.7	12.8	EN	Tier 2-4	Districts
2 months	Traffic Stops	30.9	17.4	EN	Tier 2-4	Districts
3 months	Traffic Stops	32.0	21.1	ASCM	Tier 2-4	Districts
6 months	Traffic Stops	35.2	22.4	DD	Tier 2	Beats
11 months	Traffic Stops	54.9	22.9	DD	Tier 2	Beats
1 month	Uses Of Force	23.4	12.2	EN	Tier 2	Districts
2 months	Uses Of Force	25.3	12.7	EN	Tier 2	Districts
3 months	Uses Of Force	25.9	13.3	EN	Tier 2	Districts
6 months	Uses Of Force	25.7	14.1	EN	Tier 2	Districts
11 months	Uses Of Force	24.6	12.6	EN-M	Tier 2	Districts

Note: Data from the Chicago Police Department.

Table 4(c): Backdating exercise results: individual Tier 1 districts, medium-run

		Prediction error (%)		Best-performing specification		
District	Outcome	Maximum	Minimum	Estimator	Donor pool	Donor type
6	Shooting Victims	42.0	25.7	DD	Tier 2	Beats
7	Shooting Victims	61.6	31.2	EN-M	Tier 2-4	Districts
9	Shooting Victims	51.5	40.5	EN	Tier 2	Beats
10	Shooting Victims	55.2	42.5	EN-M	Tier 2-4	Beats
11	Shooting Victims	65.0	42.9	EN-M	Tier 2	Districts
15	Shooting Victims	52.4	42.1	EN-M	Tier 2-4	Beats
6	Violent Felonies	18.1	6.7	DD	Tier 2-4	Beats
7	Violent Felonies	35.1	9.4	EN-M	Tier 2-4	Districts
9	Violent Felonies	23.3	13.8	EN-M	Tier 2-4	Districts
10	Violent Felonies	21.0	11.2	EN-M	Tier 2-4	Districts
11	Violent Felonies	35.1	11.8	EN-M	Tier 2-4	Districts
15	Violent Felonies	23.8	17.9	EN-M	Tier 2-4	Beats
6	Part 1 Crimes	14.0	3.7	DD	Tier 2	Beats
7	Part 1 Crimes	24.5	8.6	ASCM	Tier 2	Districts
9	Part 1 Crimes	15.2	6.0	EN	Tier 2	Districts
10	Part 1 Crimes	13.3	9.2	DD	Tier 2-4	Districts
11	Part 1 Crimes	16.1	7.2	EN-M	Tier 2-4	Districts
15	Part 1 Crimes	12.3	6.9	DD	Tier 2-4	Beats
6	All Crimes	10.1	3.4	DD	Tier 2	Beats
7	All Crimes	33.0	7.1	EN-M	Tier 2-4	Districts
9	All Crimes	8.3	3.5	EN	Tier 2	Districts
10	All Crimes	8.7	4.8	DD	Tier 2-4	Districts
11	All Crimes	37.0	4.5	EN-M	Tier 2	Districts
15	All Crimes	12.2	5.2	EN-M	Tier 2	Districts

Table 4(c): Backdating exercise results: individual Tier 1 districts, medium-run (cont.)

		Prediction error (%)		Best-performing specification		
District	Outcome	Maximum	Minimum	Estimator	Donor pool	Donor type
6	Gun Arrests	56.8	36.8	DD	Tier 2	Districts
7	Gun Arrests	50.4	31.5	DD	Tier 2	Districts
9	Gun Arrests	52.2	35.7	ASCM	Tier 2	Districts
10	Gun Arrests	46.9	36.6	DD	Tier 2-4	Beats
11	Gun Arrests	58.0	47.9	EN-M	Tier 2-4	Districts
15	Gun Arrests	42.8	36.1	EN	Tier 2	Districts
6	Warrant Arrests	26.5	18.8	EN-M	Tier 2	Beats
7	Warrant Arrests	44.3	17.3	EN-M	Tier 2-4	Districts
9	Warrant Arrests	31.9	22.1	DD	Tier 2-4	Beats
10	Warrant Arrests	33.7	20.2	SCM	Tier 2-4	Beats
11	Warrant Arrests	52.4	26.2	EN-M	Tier 2-4	Districts
15	Warrant Arrests	32.2	17.3	EN-M	Tier 2	Districts
6	Drug Arrests	58.3	19.8	SCM	Tier 2-4	Beats
7	Drug Arrests	86.6	26.5	EN-M	Tier 2-4	Districts
9	Drug Arrests	61.9	16.4	EN-M	Tier 2-4	Beats
10	Drug Arrests	70.1	27.7	EN-M	Tier 2	Districts
11	Drug Arrests	85.2	16.7	EN	Tier 2	Districts
15	Drug Arrests	80.5	14.9	EN-M	Tier 2	Districts
6	Misdemeanor Arrests	20.6	9.4	DD	Tier 2	Districts
7	Misdemeanor Arrests	40.3	14.6	EN-M	Tier 2-4	Districts
9	Misdemeanor Arrests	24.5	15.2	DD	Tier 2	Districts
10	Misdemeanor Arrests	19.3	11.8	DD	Tier 2-4	Beats
11	Misdemeanor Arrests	27.1	11.8	EN-M	Tier 2	Districts
15	Misdemeanor Arrests	36.9	15.1	EN-M	Tier 2	Districts
6	All Arrests	18.4	9.7	EN	Tier 2-4	Districts
7	All Arrests	43.7	11.9	EN-M	Tier 2-4	Districts
9	All Arrests	21.9	11.5	SCM	Tier 2	Districts
10	All Arrests	22.9	11.3	EN-M	Tier 2	Districts
11	All Arrests	48.4	12.2	EN	Tier 2	Districts
15	All Arrests	40.0	15.8	EN-M	Tier 2	Districts

Table 4(c): Backdating exercise results: individual Tier 1 districts, medium-run (cont.)

		Prediction error (%)		Best-performing specification		
District	Outcome	Maximum	Minimum	Estimator	Donor pool	Donor type
6	Traffic Stops	72.9	44.1	SCM	Tier 2-4	Districts
7	Traffic Stops	87.2	46.1	SCM	Tier 2	Beats
9	Traffic Stops	71.6	21.8	SCM	Tier 2	Beats
10	Traffic Stops	72.3	18.8	DD	Tier 2-4	Districts
11	Traffic Stops	91.7	53.0	DD	Tier 2	Beats
15	Traffic Stops	52.3	43.2	DD	Tier 2	Districts
6	Uses Of Force	47.2	21.7	DD	Tier 2	Districts
7	Uses Of Force	53.3	30.9	EN-M	Tier 2	Districts
9	Uses Of Force	54.7	36.9	ASCM	Tier 2	Districts
10	Uses Of Force	41.9	28.2	EN	Tier 2-4	Districts
11	Uses Of Force	59.8	19.0	EN	Tier 2	Districts
15	Uses Of Force	45.4	22.9	EN-M	Tier 2	Districts

Note: Data from the Chicago Police Department.

Table 5: Estimated effects of the SDSCs: individual Tier 1 districts, medium-run

District	Outcome	Rate per 100K (post-treatment) Actual	Counterfactual	Difference Units	%	*p*-value w/o Resampling	w/ Resampling	*q*-value
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
6	Shooting Victims	22.5	20.9	14.6	8.0	0.500	0.379	0.639
7	Shooting Victims	31.6	48.5	-113.0	-34.8	0.059	0.004	0.023
9	Shooting Victims	10.6	10.4	2.9	1.7	0.500	0.587	0.886
10	Shooting Victims	22.9	18.0	55.8	27.2	0.294	0.292	0.639
11	Shooting Victims	47.6	47.3	2.7	0.7	1.000	0.960	1.000
15	Shooting Victims	37.6	33.5	29.8	12.2	0.235	0.130	0.481
6	Violent Felonies	199.7	195.1	40.4	2.4	0.941	0.893	0.961
7	Violent Felonies	250.4	283.6	-222.6	-11.7	0.118	0.076	0.198
9	Violent Felonies	78.5	95.9	-285.5	-18.1	0.059	0.012	0.093
10	Violent Felonies	121.2	141.5	-231.1	-14.3	0.059	0.015	0.093
11	Violent Felonies	278.2	288.8	-83.9	-3.7	0.647	0.577	0.798
15	Violent Felonies	167.7	169.5	-13.3	-1.1	0.941	0.871	0.961
6	Part 1 Crimes	638.4	630.6	68.2	1.2	0.750	0.729	0.838
7	Part 1 Crimes	651.2	734.7	-559.6	-11.4	0.125	0.031	0.134
9	Part 1 Crimes	253.6	271.8	-299.1	-6.7	0.250	0.120	0.242
10	Part 1 Crimes	334.3	358.7	-278.0	-6.8	0.588	0.442	0.589
11	Part 1 Crimes	646.5	680.8	-272.1	-5.0	0.412	0.370	0.587
15	Part 1 Crimes	403.4	481.9	-573.8	-16.3	0.235	0.104	0.239
6	All Crimes	1600.0	1552.0	421.7	3.1	0.375	0.245	0.375
7	All Crimes	1908.9	2017.4	-727.3	-5.4	0.529	0.444	0.589
9	All Crimes	606.8	652.3	-748.7	-7.0	0.250	0.043	0.148
10	All Crimes	938.4	957.8	-220.8	-2.0	0.706	0.719	0.838
11	All Crimes	2087.5	2127.1	-314.0	-1.9	0.750	0.621	0.798
15	All Crimes	1184.7	1334.5	-1095.2	-11.2	0.125	0.003	0.055

Table 5: Estimated effects of the SDSCs: individual Tier 1 districts, medium-run (cont.)

District	Outcome	Rate per 100K (post-treatment) Actual	Counterfactual	Difference Units	%	*p*-value w/o Resampling	w/ Resampling	*q*-value
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
6	Gun Arrests	27.9	22.2	49.5	25.3	0.250	0.083	0.301
7	Gun Arrests	44.5	35.9	57.7	24.0	0.250	0.131	0.315
9	Gun Arrests	5.4	8.5	-52.2	-37.2	0.125	0.026	0.153
10	Gun Arrests	23.5	12.8	121.8	83.3	0.059	0.001	0.023
11	Gun Arrests	32.3	22.4	78.6	44.3	0.059	0.049	0.224
15	Gun Arrests	15.9	17.3	-10.6	-8.4	0.625	0.513	0.520
6	Warrant Arrests	55.3	47.7	67.1	16.0	0.250	0.060	0.250
7	Warrant Arrests	123.0	96.2	179.2	27.8	0.059	0.002	0.025
9	Warrant Arrests	25.2	22.2	49.7	13.7	0.471	0.434	0.460
10	Warrant Arrests	35.3	32.8	29.4	7.9	0.353	0.222	0.424
11	Warrant Arrests	112.0	100.1	94.9	12.0	0.294	0.281	0.459
15	Warrant Arrests	56.1	63.7	-55.4	-11.9	0.375	0.322	0.459
6	Drug Arrests	38.3	25.0	117.4	53.4	0.118	0.012	0.087
7	Drug Arrests	59.1	38.0	141.3	55.5	0.059	0.001	0.023
9	Drug Arrests	12.5	14.0	-25.4	-11.0	0.353	0.104	0.315
10	Drug Arrests	71.3	27.2	503.1	162.3	0.125	0.003	0.025
11	Drug Arrests	196.4	189.4	55.5	3.7	1.000	0.886	0.681
15	Drug Arrests	60.3	66.1	-41.9	-8.7	1.000	0.821	0.667
6	Misdemeanor Arrests	255.9	245.8	89.2	4.1	0.750	0.670	0.603
7	Misdemeanor Arrests	297.2	312.9	-104.8	-5.0	0.706	0.654	0.603
9	Misdemeanor Arrests	79.3	73.1	101.4	8.4	0.750	0.650	0.603
10	Misdemeanor Arrests	144.6	149.2	-53.3	-3.1	0.824	0.911	0.681
11	Misdemeanor Arrests	307.1	247.4	473.9	24.1	0.250	0.048	0.224
15	Misdemeanor Arrests	157.7	205.7	-351.0	-23.3	0.250	0.125	0.315
6	All Arrests	449.5	420.1	258.2	7.0	0.529	0.365	0.459
7	All Arrests	610.0	554.2	373.8	10.1	0.294	0.286	0.459
9	All Arrests	147.9	156.8	-145.5	-5.6	0.500	0.379	0.459
10	All Arrests	343.6	275.7	775.5	24.7	0.250	0.151	0.342
11	All Arrests	766.1	666.6	789.6	14.9	0.375	0.123	0.315
15	All Arrests	331.0	399.0	-496.7	-17.0	0.375	0.268	0.459

Table 5: Estimated effects of the SDSCs: individual Tier 1 districts, medium-run (cont.)

District	Outcome	Rate per 100K (post-treatment) Actual	Counterfactual	Difference Units	%	p-value w/o Resampling	w/ Resampling	q-value
		(1)	(2)	(3)	(4)	(5)	(6)	(7)
6	Traffic Stops	1804.2	930.9	7675.1	93.8	0.176	0.124	0.315
7	Traffic Stops	3909.7	3008.1	6042.7	30.0	0.375	0.217	0.424
9	Traffic Stops	547.4	771.2	-3680.0	-29.0	0.375	0.188	0.393
10	Traffic Stops	666.9	819.0	-1734.9	-18.6	0.412	0.395	0.459
11	Traffic Stops	2426.6	1349.8	8545.9	79.8	0.250	0.182	0.393
15	Traffic Stops	1748.8	1094.5	4783.3	59.8	0.375	0.293	0.459
6	Uses Of Force	15.9	22.3	-56.1	-28.6	0.125	0.006	0.036
7	Uses Of Force	36.0	29.6	42.4	21.3	0.250	0.114	0.265
9	Uses Of Force	6.0	5.6	6.5	7.0	0.750	0.661	0.494
10	Uses Of Force	11.9	13.2	-14.1	-9.4	0.647	0.592	0.494
11	Uses Of Force	47.3	40.7	51.7	16.0	0.250	0.125	0.265
15	Uses Of Force	18.6	20.0	-9.9	-6.8	0.750	0.621	0.494

Note: Data from the Chicago Police Department. Table reports estimated effects of the SDSCs in each Tier 1 district on crime and police activity outcomes from SDSC launch (February 2017 for the 7$^{\text{th}}$ and 11$^{\text{th}}$ districts, and March 2017 for the remaining Tier 1 districts) through December 2017. Sharpened two-stage q-values controlling the FDR calculated using the procedure described in Benjamini et al. (2006) and implemented by Anderson (2008).

A Modified Elastic Net Estimator

The elastic net (EN) estimator introduced by Doudchenko and Imbens (2017) chooses both a vector of weights and an intercept to solve:

$$(\boldsymbol{w}^*, \mu^*) = \arg\min_{\boldsymbol{w},\mu} \sum_{t=1}^{T_0} \left(Y_{0t} - \mu - \sum_{j=1}^{J} w_j Y_{jt} \right)^2 + \lambda \left(\frac{1-\alpha}{2} \|\boldsymbol{w}\|_2^2 + \alpha \|\boldsymbol{w}\|_1 \right)$$

The EN estimator imposes no direct constraint on the weights, which can be negative and sum to any value. Instead, their value is indirectly constrained by an elastic net penalty term that prioritizes weight vectors with fewer non-zero ($\|\boldsymbol{w}\|_1 = \sum |w_j|$) and smaller ($\|\boldsymbol{w}\|_2^2 = \sum w_j^2$) entries.

The magnitude of the elastic net penalty, and the relative weight placed on the Lasso and ridge regression terms, are determined by a pair of hyperparameters (α^*, λ^*). The EN estimator chooses these through a cross-validation procedure:

1. For a given pair of (α, λ), estimate a weight vector and intercept for each donor unit j: $(\hat{\boldsymbol{w}}(j; \alpha, \lambda), \hat{\mu}(j; \alpha, \lambda))$.

2. Using the estimated weight vector and intercept, calculate the predicted post-period ($t > T_0$) outcome series for each donor unit: $\hat{Y}_{jt}(\hat{\boldsymbol{w}}, \hat{\mu}) = \hat{\mu}(j; \alpha, \lambda) + \sum_{i \neq j}^{J} \hat{w}_i(\alpha, \lambda) Y_{it}$.

3. Finally, calculate the cross-validation error across all donor units as the average difference between the observed and predicted post-period outcome series: $\text{CV}(\alpha, \lambda) = \frac{1}{J} \sum_{j=1}^{J} (Y_{jt} - \hat{Y}_{jt})^2$.

4. Repeat steps (1)-(3) for different values of (α, λ), where: $(\alpha^*, \lambda^*) = \arg\min_{\alpha,\lambda} \text{CV}(\alpha, \lambda)$

This procedure chooses hyperparameters that minimize error in the post-period among the donor units, under the assumption that, because they are untreated, their post-period error should be zero. This procedure yields a single pair of hyperparameters to be used in determining weights for all the treated units, relying only on data from the donor units. This may be particularly problematic in our setting, where the six Tier 1 districts are outliers relative to the donors. The hyperparameters minimize post-period error among those donor units may not yield the most reliable comparisons for the treated units.

We propose a modification to the EN estimator that aims to address these concerns. This modification departs from the original EN estimator by choosing hyperparameters *separately* for each treated unit using a time-series cross-validation technique:

1. For each treated unit, use a subset of sequential pre-periods as a training set, with the one pre-period following as a test set.

2. For a given value of (α, λ), estimate a weight vector and intercept using the training set, and use these to predict the outcome in the single-period test set, recording the MSE.

3. Then, lengthen the training set by one period, using the next available pre-period as the new test set.

4. Repeat this process until the last pre-period is the test set.

5. Calculate the cross-validation error as the average MSE across all test sets. The values of (α^*, λ^*) minimize this cross-validation error.

Unlike the original EN estimator, which chooses hyperparameters that minimize error in the *post-period* among the donor units, the modified version chooses hyperparameters that minimize error in the *pre-period* for just the treated unit in question. The implicit assumption in the modified approach (that treatment has not yet occurred in the pre-period) is considerably weaker than the one in the original Doudchenko and Imbens (2017) approach (that the donor units are unaffected by treatment in the post-period). Furthermore, by choosing hyperparameters that minimize prediction error within the treated unit across a series of one-step-ahead forecasts, the estimator builds in some additional protection against overfitting.

B Additional SDSC Impact Estimation Results

B.1 Backdating exercise

The data-driven procedure we use to choose a model specification that minimizes out-of-sample prediction error for each district and outcome is a backdating exercise suggested by Abadie (2021). We start with incident-level data on shooting victimizations, arrests, and uses of force in Chicago from January 2013 through the end of January 2016, just before the first SDSCs launched in February 2017 in the 7$^{\text{th}}$ and 11$^{\text{th}}$ districts. We aggregate these data to either the beat-month or district-month level.[77] Then, for a given analysis type, treated unit, and outcome, such as shootings in the 7$^{\text{th}}$ district in the short-run (February 2017), we do the following:

1. Choose a model specification, consisting of an estimator (DD, SCM, ASCM, EN, EN-M), donor pool (donor units in the same or adjacent Tier, all donor units), and donor unit type (districts, beats). For example, for the analysis of short-run impacts in the 7$^{\text{th}}$ and 11$^{\text{th}}$ districts described in section 4.2, one candidate donor pool includes only the four other as-yet-untreated Tier 1 districts, while another candidate donor pool may include all the as-yet-untreated districts in the city.

2. Using the chosen specification, construct a comparison unit using at least the first $T_0^{\text{placebo}} \geq \underline{T_0}^{\text{placebo}}$ months of pre-treatment data ("training data"), where $\underline{T_0}^{\text{placebo}} = 26$.[78]

3. Compute $\hat{Y}_{0,t}(0)$ in the last $T_0 - T_0^{\text{placebo}}$ months of the pre-treatment period ("validation data").

4. Calculate an RMSPE measuring the deviation between $\hat{Y}_{0,t}(0)$ and $Y_{0,t}^{obs}$ in the first w months of the validation data, where w is the width of the outcome window for the given analysis. For example, for the analysis of short-run impacts in the 7$^{\text{th}}$ and 11$^{\text{th}}$ districts described in section 4.2, where the outcome period is February 2017, $w = 1$.

5. Repeat (2)-(4) for increasing values of T_0^{placebo}, such that the width of the hold out data is at least as large as the width of the outcome window ($T_0 - T_0^{\text{placebo}} > w$).

6. Repeat (5) for each model specification.

[77] Reliable incident-level data for all of our outcomes exists starting in 2010. However, due to changes in the boundaries of districts that occurred before 2013, we use only data from 2013 onward when the definitions of Chicago's police districts remained stable.

[78] For discussion of the sensitivity of the backdating exercise to the choice of $\underline{T_0}^{\text{placebo}}$, see section 4.2 of the main text and Figure 7.

This procedure yields for each analysis type, treated unit, outcome, and model specification a series of measures of out-of-sample prediction error (RMSPEs), one for each value of T_0^{placebo}. We average over these RMSPEs and choose the specification that minimizes this average error. Rather than relying on theory to select an estimator or donor pool composition, this procedure uses pre-treatment data to guide our choice, and does so separately for each analysis type, district, and outcome.

The results of this exercise are reported in Appendix Tables 4(a), 4(b), and 4(c), and visualized in Appendix Figure 14.

Several findings are noteworthy:

- First, there is significant variation across model specifications in average out-of-sample prediction error. For example, for shootings in the 7^{th} district in the short-run analysis, the worst-performing specification yields an average RMSPE that is 54% of the true rate of shootings, compared to 27.5% for the best-performing specification, which relies on the EN-M estimator and a donor pool that includes all other districts.
- Second, there is no single model specification that consistently outperforms all others across outcomes, treated units, and analysis types.
- Third, average prediction error varies widely across outcomes, with shootings, traffic stops, and uses of force being the hardest for these models to predict, and arrests being the easiest, with the exception of drug arrests and, to a lesser extent, gun arrests.

These results also confirm our concerns about the reliability of the estimates for individual Tier 1 districts (labeled "medium-run" analysis in Appendix Figure 14). The minimum prediction errors within each outcome tend to be higher, often considerably so, for this analysis compared to the short-run analysis focused only on the 7^{th} and 11^{th} districts (using a broader donor pool) or the analysis of the Tier 1 districts as a whole (using the same narrower donor pool). For these reasons, we focus our attention in the main text on the short-run analysis of the 7^{th} and 11^{th} districts and the medium-run analysis of the Tier 1 districts as a whole, but report results from the medium-analysis of individual Tier 1 districts in Appendix B.2.

B.2 Medium-run analysis for individual Tier 1 districts

Due to the higher prediction error in the medium-run analysis for individual Tier 1 districts, we focus in the main text on results from the short-run analysis of the 7^{th} and 11^{th} districts and the medium-run analysis of the Tier 1 districts as a whole. For completeness, we report results from the medium-run analysis for individual Tier 1 districts in Appendix Table 5.

The SDSC in the 7th district is estimated to have reduced the rate of shooting victimization by almost 35%, resulting in 113 fewer victims, from February to December 2017. This estimate is below conventional thresholds of statistical significance, both before adjusting for multiple comparisons ($p < 0.01$) and after ($q < 0.05$). However, in every other Tier 1 district, the SDSCs' estimated effects on shooting victimization range from close to zero (11th and 9th districts), to modest increases (6th and 15th districts), to a large increase in the 10th district, though none of these estimates are statistically significant either before or after adjusting for multiple comparisons. The SDSCs' are also estimated to have reduced violent felonies by between 14% and 18% in the 9th and 10th districts ($p < 0.05$, $q < 0.1$), with mostly negative but imprecise estimates in other districts.

The SDSCs' estimated effects on overall arrests range from a 17% decrease (15th district) to a 25% increase (10th district), though none of these estimates are statistically distinguishable from zero. However, consistent with the patterns reported in section 4.3, we see some evidence of the SDSCs affecting the composition of arrests, with positive effects on drug arrests in the 6th, 7th, and 10th districts ($p < 0.05$, $q < 0.1$), gun arrests in the 10th district ($p < 0.01$, $q < 0.05$), and warrant arrests in the 7th district ($p < 0.01$, $q < 0.05$). Finally, the SDSCs' estimated effects on uses of force are similarly heterogeneous and imprecise, with the exceptions of the 6th district, which is estimated to have experienced a 29% reduction ($p < 0.01$ and $q < 0.05$), or 56 fewer uses of force.

Some of this heterogeneity may reflect variation in the SDSCs' true effects on outcomes, including due to variation in how the SDSCs were implemented, a point we discuss in section 5. Some of it, however, may also reflect estimation error, owing to the very difficult task of constructing comparison districts from the much more limited donor pool in this analysis. This estimation error is likely to be most acute for the West side SDSC districts (10th, 11th, 15th) since there were almost no high-violence areas on the West side left untreated at that point, save for a few beats bordering them in adjacent districts; as reported in Appendix Table 4(c), out-of-sample prediction errors for shootings are generally higher in these three districts than in those on the South side.[79]

[79] The exception to this is the 9th district, which has a high prediction error for shootings and is on the South side. The 9th district, along with the 10th, both have a high shooting prediction error and are unusually racially heterogeneous relative to the other Tier 1 districts; in the 9th, the population is approximately 10% African-American and 60% Hispanic, and in the 10th, it is roughly one third African-American and two-thirds Hispanic. To the extent that patterns of violence may differ across predominantly African-American and Hispanic neighborhoods, the 9th and 10th may be among the more challenging districts for which to construct a comparison. In contrast, all the other Tier 1 districts are more racially homogeneous, having populations that are 80% or more African-American. The 6th and 7th districts, with the lowest shooting prediction errors, have populations that are 97 and 93% African-American, respectively.

C Potential Conflicts of Interest

Our research center (the University of Chicago Crime Lab) was involved in the delivery of the intervention described in sections 3, 4, and 5 of this paper. This creates natural concerns about the potential for a conflict of interest. In this Appendix, we provide additional details on the nature of our involvement, followed by how we deal with conflict of interest concerns.

C.1 Crime Lab involvement with the SDSCs

In 2016, Chicago was struggling with two different crises within law enforcement. The first crisis was the crisis of public confidence and trust following the murder of 17-year-old Laquan McDonald, who was shot 16 times by CPD officer Jason Van Dyke on October 20, 2014, and the public release of the dashboard camera footage of his death on November 24, 2015. The second was a 60% surge in murders in from 2015 to 2016. In response to the first crisis, the city launched a Police Accountability Task Force, while the U.S. Department of Justice (DOJ) Civil Rights Division launched an investigation of CPD. As part of the response to the second crisis, the rise in violent crime, Chief Sean Malinowski of LAPD was invited in late 2016 to conduct an assessment of CPD's anti-violence strategies and their implementation. The idea for the design of the SDSCs is due to Malinowski, and the decision to implement them was made by the City of Chicago.

The Crime Lab's involvement arose from the City's desire to implement the SDSCs as quickly as possible given the public safety and public health consequences of the rise in gun violence, especially in the Chicago's most economically and socially vulnerable communities. As described in section 3, a key part of the SDSC model is the role of the civilian crime analyst. However, no personnel at CPD or other City agencies were available and qualified to serve in that role for the launch of the first two SDSCs in the 7$^{\text{th}}$ (Englewood) and 11$^{\text{th}}$ (Garfield Park) police districts in February 2017. Moreover, the City's hiring process made it unlikely that any qualified applicants could be hired within an acceptable timeframe. The Crime Lab was set up to use data and data science to help cities reduce gun violence and the harms of the administration of criminal justice itself by partnering with public sector agencies. Because it had already developed a productive working relationship with many parts of City government, the Crime Lab was asked to help fill this short-term staffing need.

In response to this request, two Crime Lab staff members were initially embedded within the 7$^{\text{th}}$ and 11$^{\text{th}}$ districts in February 2017 as civilian crime analysts. When the City expanded the number of SDSCs—as they did in March 2017 (to four more Tier 1 districts), January 2018 (to four Tier 2 districts), and March 2018 (to three more Tier 2 districts)—Crime Lab staff were embedded as analysts in these additional districts as well, sometimes with one Crime Lab analyst assisting multiple districts. The crime analyst's role was to assemble and summarize statistical information about patterns of criminal activity for district commanders and their staff. The analyst's role was not to recommend specific missions or officer activities, nor to make policy decisions. By October 2018, CPD hired 13 of its own civilian crime analysts, one for each of the 13 Tier 1 and Tier 2 districts. As a result, the role of Crime Lab staff transitioned from providing direct analytical support to SDSC districts to developing training materials for future analysts and working on other technical assistance projects with CPD.

If the Crime Lab was involved with the delivery of the SDSC intervention, why would we also carry out a study of that intervention? After all, an independent evaluation of the SDSCs was conducted by the RAND Corporation (Hollywood et al., 2019) at the request of the Bureau of Justice Assistance at the DOJ and CPD. There are three reasons. First, although implementing and evaluating an intervention creates natural concerns about the potential for a conflict of interest, doing so also creates complementarities from developing a deeper understanding of the intervention being studied, as well as economies of scale. For example, it was through gaining familiarity with the HunchLab predictive policing software as part of the implementation work that we came both to understand its potential usefulness as a test of SDSC implementation fidelity across districts and to access the data necessary to operationalize that test. A second reason for us to carry out this study is that we think the conceptual issues raised by the SDSCs about the nature of policing, and how to improve it, extend far beyond a narrow examination of the impacts of the SDSCs themselves, as reflected in the broader set of analyses we carry out in this paper. Finally, our research team has a different perspective about the best methodological approach to measuring the SDSCs' impacts. As we discuss in the main text (section 4), constructing suitable comparison units for SDSC districts is challenging because they are so different from the other districts in the city, both in terms of their historical levels of violence and the size of the surge in violence they experienced in 2016, immediately prior to the launch of the initial SDSCs. The two-way fixed-effects difference-in-differences estimator used by (Hollywood et al., 2019) struggles to construct comparison units that maximize out-of-sample predictive accuracy, as shown by the results of the backdating exercise we conduct (see Appendix B.1). The other potential concern with (Hollywood et al., 2019) is that recent work has shown, in settings such as this one with staggered treatment adoption and likely heterogeneous treatment effects, that two-way fixed effects difference-in-differences estimates can exhibit substantial bias (Goodman-Bacon, 2021).

The Crime Lab's funding model is not that of a 'fee for service' organization; that is, the funding for most Crime Lab research projects comes not from the public sector agencies or non-profit organizations implementing the policies or interventions we study, but instead from third parties such as philanthropic foundations; federal research funders such as NIH, NSF, or the U.S. Departments of Education or Justice; or individual philanthropists. But with the SDSCs, the need to deploy a large number of Crime Lab staff quickly limited the time available to raise funds from third party sources to cover those costs. This exceptional case was the one time the Crime Lab agreed to accept funding from the City of Chicago directly for its work. Funding for other SDSC costs (such as the costs of the technology build-out), eventually extending delivery in later years, and our evaluation work was provided by the John D. and Catharine T. MacArthur Foundation, Ken Griffin, and the University of Wisconsin's Institute for Research on Poverty.

Therefore, the two most worrisome sources of a potential conflict of interest in this work are that: (1) we are studying an intervention that, while not our idea, we contributed to implementing; and (2) we are studying an intervention that was carried out by the City of Chicago, and the City contributed to covering some of our costs.

C.2 Managing potential conflicts of interest

While there does not appear to be a consensus on the best way to manage potential conflicts of interest in research (as evidenced by ongoing debates in areas such as medicine), there is little disagreement (including from us) that additional steps beyond standard research practice are imperative. The approach we take is guided by two principles: (1) accountability and (2) transparency.

For the goal of *accountability*, we are committed to having all of our work subject to rigorous external scientific peer review (recognizing that within economics there is also a tradition of releasing working papers for discussion purposes while peer review is underway). For peer review in economics, the editor of the journal to which the paper has been submitted (a leading researcher) selects on their own a set of outside scientific experts who read and critique the paper and provide anonymous comments to the study authors, suggestions for changes, and a recommendation to the editor about whether the paper should be published, rejected, or whether the authors should be offered the chance to revise and resubmit their paper for re-consideration. In principle, another step we could have taken at the outset of this research was to file a pre-analysis plan that described what analytical decisions we would make before knowing any results. Pre-analysis plans can work well when the analytic strategy and key outcome of scientific interest are fairly straightforward, as is often the case in randomized controlled trials. However, it is less clear how beneficial they are in a setting like this one where both the research design and key outcomes are ex ante much less obvious. For more discussion on the complexities of pre-analysis plans, see Olken (2015).

For the goal of *transparency*, as a complement to scientific peer review, we adopt two different, related strategies. For starters, we show as many sensitivity analyses as possible about how our results do or do not change in response to different reasonable decisions about the estimation procedure. We are also committed to making as much of our data publicly available as we can, subject to data confidentiality requirements.[80] We will publicly disclose, on the University of Chicago Crime Lab's GitHub page,[81] two datasets used in this paper. The first dataset contains counts, at the police beat-month level, of selected crimes, arrests, and measures of police activity from 2010 through 2019, enabling others to replicate the analysis in section 4. The second dataset contains, at the HunchLab box-day-watch level, information on a box's predicted risk, whether it was shown or not, ShotSpotter alerts, and officer time, enabling others to replicate our analysis in section 5.

[80] For example, we cannot disclose data at the level of the individual reported crime or arrest that might reveal the identities of individual people involved with such incidents.

[81] https://github.com/uchicago-urbanlabs-crimelab