Supplementary Information

Supplementary Methods

Test sidedness

Coefficient tests used two-sided normal or t references as specified for each design. Holiday bootstrap P values doubled the smaller tail probability. Marathon matched-date permutation tests compared absolute observed and permuted contrasts, counting departures in either direction. The circular-shift pulse-shape test instead used the upper tail of the mean subgroup-to-city correlation distribution: it tested greater positive alignment and was one-sided. Moran's I permutation tests (999 permutations) likewise used the upper tail, testing positive spatial autocorrelation. The within-borough permutation test of the early-month race slope compared absolute permuted and observed slopes and was two-sided. All four permutation procedures included the observed configuration in the numerator and denominator adjustment. Joint Wald and Cochran heterogeneity tests used upper-tail chi-squared references; joint tests across income quartiles used upper-tail F references with G-1 denominator degrees of freedom. Calendar-validation direction checks were descriptive, not additional significance tests.

Timestamp handling and endpoint constructs

The public field cannot identify which physical occurrence generated a record during the repeated fall-back 01:00 hour. The parser assigns those values to the earlier fold and flags them; 6,128 of 17,490,489 raw study-period rows (0.035%) were ambiguous. Day-level calendar assignment is unchanged by the fold, and the primary holiday windows and active marathon window do not overlap it. The ZIP-hour panel uses a continuous UTC-hour index, and hourly NLDAS-2 weather was linked by ZIP code and UTC hour. The exact raw snapshot, parser and ambiguity audit are identified in Data and Code availability, and the full timestamp-provenance and daylight-saving audit is reported in Supplementary Table S20.

A construct sensitivity used INITIAL_CALL_TYPE and separately removed VENOM; 410 of 1,172,858 raw study-period records in the final DRUG*/VENOM category were VENOM. A separate full-raw-file construct audit repeated the early-month count and share models using initial DRUG*/VENOM, final DRUG*/VENOM, and final DRUG* with VENOM excluded; it did not replace the weather-panel primary model (Extended Data Table 3).

ZIP income quartiles

Income quartiles used American Community Survey 2015-2019 median household income. Cut points came from the 182 ZIP Code Tabulation Areas with land in the five NYC counties and eligible income estimates, with postal ZIP 11249 assigned to ZCTA 11211. The 170 study ZIPs fell into quartiles of 46, 44, 46 and 33 ZIPs (one without an income estimate); the marathon design used the same assignments.

Temporal validation and endpoint mapping

The locked calendar-event response-time classifier used total-call-weighted city-hour means to reproduce the earlier estimand. It required an early-month estimate no larger than +30 seconds and negative estimates for both celebration classes. The original combined decision also required a positive permit-linked access estimate, but that separately defined analysis did not test recurrence of the calendar contrasts. The response criteria were directional because the validation period contained only two July 4 and four pooled Memorial/Labor occurrences. Secondary summaries were the Spearman correlation across all 15 count cells, the correlation after excluding burns and year-specific point estimates. No alternative window, endpoint subset or model was used after inspection (Supplementary Table S18).

Endpoint mappings were fixed from each source vocabulary before calendar contrasts were estimated. Phoenix and Glendale used regional codes for overdose (OD), assault (ASSLT, ASSLTM, ASSLTS), burn (BURN) and cardiac arrest (CODE variants). Cincinnati used raw incident-type prefixes for drug or poisoning (23, OD, HER), assault (4, ASSLT), burn (7, BURN) and cardiac arrest (9, ARREST). The Cincinnati crosswalk was required to have at least 0.98 sensitivity and positive predictive value against populated official categories; the observed minima were 0.996 and 0.990. PDOA was excluded because Cincinnati classifies it as possible dead on arrival rather than cardiac arrest. No defensible common mental-health category was available, so that endpoint was not transported. Direct call-share models and temperature-adjusted models were diagnostics; a 2015-2022 Cincinnati sensitivity preceded the collapse of its official group field.

The complete interval was analysed first and underlies the legacy component, distance and distributional analyses. We prioritize call-to-scene because it directly measures pre-arrival access and avoids conditioning on completed transport; this outcome hierarchy was not preregistered. After identifying that mismatch, we froze a dated principal-outcome phase and event-level specification before fitting it. The phases were prior Saturday 21:00-23:59, Sunday setup 00:00-06:59, active event operations 07:00-20:59 and post-event 21:00-23:59. This definition avoids calling all pre-07:00 Sunday hours pre-closure because some route segments close earlier; the superseded early-Sunday comparison is retained for transparency in Extended Data Fig. 1.

We estimated the all-day difference-in-differences with ZIP and year-month fixed effects, hour fixed effects, a temperature spline and a precipitation indicator, with standard errors clustered by ZIP and date. The two-way

covariance matrix required positive-semidefinite repair in 57 of 67 legacy marathon models; one-way date- and ZIP-clustered sensitivities preserved the headline inference (Supplementary Table S6). The principal phase model instead included ZIP by hour-of-week and physical UTC-hour fixed effects and estimated the four phase contrasts within each held event and its four matched Sundays. We summarized the 10 annual effects equally using t inference with 9 degrees of freedom, repeated the analysis after each event omission, and permuted the event label among the five matched Sundays 19,999 times. The active and prior-Saturday phases do not overlap the ambiguous fall-back hour. The secondary complete-interval result was also robust to excluding incidents before 03:00 (Supplementary Table S6). For the off-diagonal ZIP-day count model, a common-window sensitivity removed the entire local 01:00 hour from every candidate Sunday, so event and control dates each contributed 23 physical hours without requiring a fold assignment (Supplementary Table S12).

Calendar response-time resolution check

Before inspecting ZIP-level estimates, we froze a resolution-parity test for the three 2012-2023 calendar response-time contrasts, using the same windows, controls, outcome and total-call weights at ZIP-hour resolution, ZIP-by-hour-of-week and ZIP-by-period fixed effects, and ZIP and date clustering. A full parity win required the early-month interval to lie within +/-30 seconds and both holiday estimates to be below -30 seconds with intervals excluding zero. Lacking valid-interval counts, the test addresses outcome resolution, not causal identification or interval missingness (Supplementary Table S25).

Marathon course reconstruction

The marathon course for 2013-2023 was reconstructed along its documented streets (Verrazzano-Narrows Bridge, 4th Avenue, Lafayette and Bedford Avenues, Greenpoint, the Pulaski and Queensboro Bridges, First Avenue, the Bronx loop, the Madison Avenue Bridge, Fifth Avenue and Central Park) by routing between labelled waypoints with FOSSGIS OSRM on OpenStreetMap, using straight segments for the Verrazzano-Narrows and Willis Avenue spans and removing out-and-back routing spurs. The result measured 42.5 km against the 42.2-km course; a predefined length gate of 2% rejected two earlier attempts (72.8 km with one-way driving restrictions and 50.2 km with routed bridge approaches). ZIP classes used 2020 ZCTA polygon centroids in UTM zone 18N, as before; recomputing distances to the previously used straight-line approximation reproduced its classification exactly, and the reconstructed course moved 19 of 248 ZIPs between classes, increasing route-adjacent ZIPs from 24 to 34. All marathon analyses were rerun with the reconstructed course.

Population presence

Calls are counted where incidents occur, so calendar patterns could reflect who is present rather than incidence. Presence was measured with Advan Weekly Patterns Plus, which records visits by a panel of mobile devices to commercial points of interest; daily visits were aggregated to the 170 panel ZIPs from 7 January 2019 to 31 December 2023 and expressed as visits per point of interest, which is robust to changes in the number of tracked places. The design was fixed before fitting. The early-month matched design and the joint calendar specification were applied to log presence, and the early-month and holiday call models were refitted on the same 2019-2023 sample without and with log presence as a covariate; a ZIP-level version used ZIP presence. July 2022, when the tracked-place panel dipped, was excluded, with sensitivities including it or dropping 2020-2021. Advan dates were checked against holiday minima. Advan data are licensed through Dewey and are not redistributed; Supplementary Table S38 reports all estimates.

Supporting early-month clinical-code and disposition results

Secondary endpoint and disposition models further characterized the pattern. Unconscious calls rose 6.2% (95% CI 4.7-7.7%), cardiac-arrest calls 4.5% (2.4-6.6%) and seizure calls 2.1% (0.7-3.5%), whereas assault and mental-health calls did not change (Extended Data Table 3). Matched incident-level models detected no association with highest-acuity dispatch (OR 0.942, 95% CI 0.843-1.053) or transport (OR 1.009, 95% CI 0.994-1.023). The dead-on-arrival estimate was low-information (OR 0.68, 95% CI 0.37-1.23), with outcome variation in only 36 matched strata.

Excess pre-arrival waiting

Total pre-arrival waiting in a cell is the number of incidents multiplied by their mean valid call-to-scene time (dispatch plus travel, more than 0 and under 3,600 seconds), so invalid intervals take the cell mean. For an exposed cell (1) and its counterfactual (0), the excess divides exactly as (N1 - N0)T0 + N0(T1 - T0) + (N1 - N0)(T1 - T0): more calls at the usual wait, the usual calls at the changed wait, and their interaction. Citywide schedules use the 170-ZIP panel, whose hourly counts rebuilt from incident records equal the published counts in every local hour; the counterfactual count is the mean over matched control days or hours and the counterfactual wait is their pooled valid mean. Street operations use a near-far counterfactual: the wait adds the distant-ZIP change to the route-adjacent control mean, and the count scales the route-adjacent control count by the distant-ZIP ratio, so the wait difference equals the published event effect. The relative change divides the excess by mean comparison waiting, using the same bootstrap draws. Annual totals multiply by 33 early-month days, one July 4, two Memorial and Labor Days, one race and five UN weekdays. Invalid intervals include both non-positive values and waits beyond the ceiling, so the cell-mean assignment can misstate the excess where validity fell, near the route (-4.9 percentage points) and in the fireworks window (-2.4). After the

first run, a code fix corrected the exposed valid fraction for multi-day UN weeks, both standby exclusions were added, and the July 4 full holiday day was estimated as a labelled diagnostic; no window, comparison or schedule changed (Supplementary Table S46).

Calendar-wide scan

To place the studied schedules within the whole calendar, we refitted the joint calendar model once per endpoint and candidate with an indicator for the candidate's dates: every calendar date outside the already-modelled holiday windows (342 dates) and 19 movable or religious observances (Super Bowl Sunday, Easter, Mother's and Father's Day, the eve of and day after Thanksgiving, Election Day, both clock-change Sundays, the Puerto Rican Day Parade and Pride March Sundays, Lunar New Year, Eid al-Fitr, Eid al-Adha, every day of Ramadan, Rosh Hashanah, Yom Kippur, the first day of Passover and Diwali). Inference used calendar-year clustering (t, 11 degrees of freedom) and Benjamini-Hochberg control within each endpoint and family; flagged candidates were refitted on 2024-2025 counts as a direction check. The design, candidate dates and inputs were hashed before fitting (Supplementary Table S47).

Supplementary Note 1. Additional results

This note reports secondary results and checks referred to in the main text. Values trace to the tables cited.

Marathon access. Equal-weight annual route-adjacent estimates remained between +133 and +164 seconds after each leave-one-event omission (Supplementary Table S23). The pooled incident-level model showed +13 seconds on the prior Saturday evening (-57 to 84; Fig. 1c; Supplementary Table S23). On the 2020 fall-back Sunday, during the pandemic and without closures, the route contrast was -2 seconds (Supplementary Table S29). The full-event-day contrast of +110 seconds (95% CI 57-163) was estimated among the 98.7% of 95,996 candidates with a valid interval (Supplementary Tables S14 and S19). Lower interval validity near the route (a differential -3.7 percentage points, 95% CI -6.0 to -1.3) did not account for the delay: assigning one second to every invalid interval still left +104 seconds (61-147; Supplementary Tables S19 and S23). After one second was assigned to all 278 invalid active-window intervals, every annual estimate remained positive (Supplementary Tables S19 and S23). Long waits also became more common. Route-adjacent incidents waiting more than 15 minutes rose by 6.8 percentage points over an 8.8% control-day base (Supplementary Table S9). About 37 additional route-adjacent incidents per marathon waited more than 15 minutes for responders, a descriptive conversion of the 6.8-percentage-point increase (Supplementary Tables S9 and S41). Among marathon calls with completed transport, call-to-hospital time was 149 seconds longer (69-229; Supplementary Tables S8 and S13). The complete call-to-hospital sample contained 65,639 records (Supplementary Table S19). The share of calls held for unit assignment rose by 5.2 percentage points (0.6-9.9). Calls never held for unit assignment were slower near the route by 114 seconds (95% CI 79-150), positive in all 10 races, although

held status can itself change with the event (Supplementary Table S37). Because every race fell in the first week of November, we checked whether the early-month pulse inflated the delay: races on days 1-3 did not show larger call-to-hospital delays than those on days 4-7 (+110 seconds, 26-194, four races; +179 seconds, 91-266, six races; Supplementary Table S13). Income-stratified marathon models did not show a monotonic gradient (Extended Data Tables 7 and 8).

United Nations weeks. Because the UN design admitted intervals up to 7,200 seconds, we refitted it under the marathon's 3,600-second ceiling; the pooled all-hours estimate fell to +74 seconds (35-113). The total-waiting synthesis, which assigns invalid intervals the mean call-to-scene time of their ZIP and hour, gives 68 seconds (Supplementary Tables S11, S23 and S46).

Passenger trips. A sensitivity using comparison-date weights gave +55 seconds (95% interval 41-70), with a matched placebo of -4 seconds (-11 to 1; Supplementary Fig. S3a and Table S28). Its common cells retained 15.9% of exposed and 28.6% of comparison trip appearances, against 23.0% and 37.7% in the primary design (Table 1; Supplementary Fig. S3b and Table S28).

Early month. Mid-month comparisons did not reproduce the early-month increase (Extended Data Table 4). Legacy Social Security advances did not show detectable retiming (Supplementary Tables S10 and S15). The early-month increase appeared in every six-hour block of the day: 4.2% overnight (00:00-05:59; 95% CI 1.9-6.7%) and 13.1% in the afternoon (12:00-17:59; 10.9-15.4%); in 2024-2025 the morning block was largest and differences between blocks were not detected (P = 0.18; Supplementary Table S30). Unconscious, cardiac-arrest and seizure calls also increased on days 1-3, including among calls never coded as drug or poisoning, whereas highest-acuity dispatch and transport showed no detected change (Supplementary Tables S12, S25 and S44). Unconscious calls increased by as much among calls never coded as drug or poisoning at intake or closure (IRR 1.061) as among all such calls (1.062), so recoding between these categories did not produce the increase. Ambiguous intakes were slightly more often closed as drug or poisoning calls on days 1-3 (ratio of proportions 1.055, 1.014-1.098), and refusals of care among drug or poisoning calls did not change (1.016, 0.994-1.038; Supplementary Table S44). Call-to-scene time showed no detected change on these days (+5.3 seconds, 95% CI -5.0 to 15.6; Supplementary Tables S12 and S25). Commercial foot traffic showed no detected rise on these days (2019-2023; Supplementary Table S38). Dispatch codes cannot show whether some of the early-month rise in unconscious and cardiac-arrest calls reflected substance-related events reported without a named substance.

Holidays. In the July 4 fireworks window, burn, assault and drug or poisoning calls also took larger shares of all calls (Fig. 3b), with adjusted odds ratios of 2.83 for burns (95% CI 2.07-3.87), 1.89 for assault (1.60-2.24) and 1.13 for drug or poisoning (1.04-1.23); the burns result rested on 167 calls. In the post-midnight segment, drug or poisoning calls rose 47% (26-71%; Supplementary Table S2). On Memorial and Labor Day, category shares

moved in the same directions as counts, partly because total calls fell, and the cardiac-arrest share rose without a detected change in its count (Fig. 3b).

Calendar-wide scan. Christmas and Thanksgiving, fitted as adjustment terms, carried signatures like those of the studied holidays, including 70.8% more burn calls on Thanksgiving (50.4-94.0) and 11.1% more cardiac-arrest calls on Christmas (5.4-17.1; Supplementary Table S47). Refits with indicators for 342 other calendar dates and 19 movable or religious observances, with false-discovery control within each endpoint, found comparable signatures only on or around a few other holidays, notably assault calls on Halloween (46.5% higher, 95% CI 30.9-63.9) and drug or poisoning calls on the eve of Thanksgiving (11.4%, 6.7-16.3), each with the same direction in 2024-2025 (Supplementary Table S47).

Overlap and neighbourhoods. Public-housing share also accompanied a larger early-month increase (+1.2% per standard deviation, 0.2-2.1; Supplementary Table S33). January's extra early-month excess came from New Year's Day; days 2-3 were no higher than in other months, and where Labor Day fell on days 1-3 the two pulses combined without a detected interaction in drug or poisoning calls (departure from additivity +3.7%, 95% CI -3.1 to 11.0; Supplementary Table S31). The racial-composition association with the early-month increase held with borough fixed effects and household shares receiving SSI, cash assistance or Supplemental Nutrition Assistance Program benefits (+3.2% per standard deviation, 2.1-4.3) and under within-borough permutation (P = 0.001; Supplementary Table S33). The same gradient appeared for unconscious calls (+3.2% per standard deviation, 2.0-4.4) and for drug or poisoning codes assigned at intake (+3.3%, 2.2-4.5; Supplementary Table S44), and early-month foot traffic rose only slightly with Black share (+0.3% per standard deviation, 0.1-0.5; Supplementary Table S38). New Year's Day ran the other way on income, rising with it. Among the other prespecified moderators, pawnbroker density did not accompany a larger early-month increase, and alcohol-licence density accompanied a larger New Year's Day excess but, contrary to the prespecified hypothesis, a smaller July 4 excess (Supplementary Table S33). In the permit census, a descriptive allocation of the pooled excess placed 53% of the excess call-seconds in Manhattan precincts, which hold 20% of residents, and 42% in the lowest-poverty third of residents against 33% in the highest-poverty third (Supplementary Table S45).

Total pre-arrival waiting. About 30% of the added marathon calls, 16% on July 4 and 44% near the UN had no valid interval and were assigned the mean call-to-scene time of their ZIP and hour (Supplementary Table S46).

Recurrence. Magnitudes were smaller in 2024-2025, at 0.61 times their 2012-2023 values across the seven prespecified cells (95% CI 0.47-0.75), although the early-month increase held fully (Supplementary Table S39). Early-month call-to-scene time remained near zero (-5.7 seconds) and the Memorial and Labor Day estimate remained negative (-202.4 seconds), but full-day July 4 reversed from -67.5 seconds in 2012-2023 to +61.7 seconds in 2024-2025, on only two later occurrences (Supplementary Table S18). Citywide holiday response

times depend on total call load, staffing and deployment, which can change from year to year even when the date does not, and these data cannot test whether that explains the reversal. The Spearman correlation between 2012-2023 and 2024-2025 log-IRRs rose from 0.914 to 0.965 after excluding burns (Supplementary Table S18). Applied without refitting to 2024-2025, holiday and day-of-month terms added 12 percentage points (95% interval 3-22) to the share of within-month day-to-day variation in drug or poisoning calls explained, about as much as realized weather (11 points), and flagged 48% of the busiest tenth of days, against 39% for the weekday-by-month rhythm alone. For total calls they added 7 points (-2 to 14), against 28 for weather (Supplementary Table S43).

External cities. Early-month drug or poisoning, July 4 burns and assault, and Memorial/Labor assault were positive in all four cities; both holiday drug or poisoning contrasts were positive in three of four. Several cells carried little information (for example, Glendale July 4 burns, 95% CI 0.03-150.7). Magnitudes were not pooled because periods and dispatch systems differ, Phoenix and Glendale share a regional code vocabulary, and Cincinnati's all-call denominator includes fire incidents. The results support spatial recurrence of dispatcher-coded count patterns, with weaker transportability for holiday drug or poisoning calls (Supplementary Table S24). In the joint calendar model, the July 4 assault excess beyond the Sunday pattern was inconsistent across cities, whereas Memorial and Labor Day call volume was at or below Sunday levels in all three cities (Supplementary Table S35).

Cincinnati marathon. The marathon design, fixed in August 2026 and applied unchanged to Cincinnati's Flying Pig Marathon in 2022-2025, gave a positive but imprecise delay among calls with a recorded arrival (+67 seconds, 95% CI -17 to 152). A validity rule added in October, before any response time was computed for six earlier races, stopped their extension: route-adjacent calls lacked a valid interval more often than distant calls by more than 10 percentage points. The 2022-2025 races, analysed before the rule existed, would also have failed it (Supplementary Table S42).

Extended Data

Extended Data Table 1. Marathon cancelled-event placebo for the complete call-to-hospital interval. Difference-in-differences applied to the completed-transport outcome for the 2012 New York City Marathon, which was cancelled in the wake of Hurricane Sandy, using the same matched-control-Sunday structure as the held-event analysis. The complete-interval delay estimated on held marathon days was not detected on the cancelled date.

Test	N	Estimate	CI_lo	CI_hi	P
Sandy 2012 cancelled-marathon placebo	6080	-85.6	-199	27.8	0.213

Extended Data Table 2. Marathon component-specific interval estimates. Separate difference-in-differences models for dispatch, travel, scene and transport start from records with a valid complete call-to-hospital

interval, then retain positive values below one hour for each component. These component-specific filters produce different samples. Travel and transport estimates were positive; the scene interval included zero, and the dispatch interval only just excluded it (lower bound 0.1 s; t-test P = 0.055). The component estimates are not additive and do not decompose the complete-interval estimate. Confidence intervals use a normal reference; P values use the model's clustered t reference.

Component	Estimate	CI_lo	CI_hi	P
dispatch	37.0	0.1	74.0	0.0553
travel	55.1	29.2	81.0	0.000123
scene	1.0	-29.7	31.6	0.95
transport	74.9	42.6	107.3	3.69e-05

Extended Data Table 3. Early-month endpoint counts, category shares and construct sensitivity. Panel A reports case-crossover count IRRs for days 1-3 versus matched later-month controls (January excluded), including the all-calls benchmark. Panel B reports binomial category-share models as adjusted odds ratios for drug or poisoning versus all other calls, comparing days 1-3 with matched controls. These are not percentage-point changes in share. Initial-code, final-code and VENOM-exclusion sensitivities use the full raw file. The construct audit uses the same calendar matching but does not impose the weather-panel ZIP restriction and does not replace the primary model.

Panel A: endpoint-specific counts and all-call benchmark

Endpoint	Model	IRR	95% CI	P
Drug or poisoning	Conditional Poisson	1.086	[1.074, 1.098]	< 1e-15
Unconscious	Conditional Poisson	1.062	[1.047, 1.077]	< 1e-15
Cardiac Arrest	Conditional Poisson	1.045	[1.024, 1.066]	2.18e-05
Assault	Conditional Poisson	1.006	[0.977, 1.035]	0.686
Mental Health	Conditional Poisson	1.001	[0.993, 1.008]	0.886
Seizures	Conditional Poisson	1.021	[1.007, 1.035]	0.0026
All calls	Conditional Poisson	1.012	[1.006, 1.018]	0.000159
Drug or poisoning relative to all calls	Delta-method log ratio	1.073	[1.060, 1.087]	< 1e-15

Panel B: direct share and endpoint-construct sensitivity

Specification	Estimand	Ratio	95% CI	P
Weather-panel final code	Share odds ratio	1.081	[1.068, 1.094]	< 1e-15
Initial DRUG* or VENOM	count IRR	1.089	[1.076, 1.103]	< 1e-15
Initial DRUG* or VENOM	share odds ratio	1.084	[1.070, 1.097]	< 1e-15
Final DRUG* or VENOM	count IRR	1.086	[1.073, 1.099]	< 1e-15
Final DRUG* or VENOM	share odds ratio	1.081	[1.068, 1.094]	< 1e-15
Final DRUG*; VENOM excluded	count IRR	1.086	[1.073, 1.099]	< 1e-15
Final DRUG*; VENOM excluded	share odds ratio	1.081	[1.068, 1.094]	< 1e-15

Extended Data Table 4. Early-month case-crossover: exact-week mid-month contrasts and year-stratified replication. Top rows: the days 1-3 versus same-weekday days 10-24 design and two placebo contrasts pairing days 15-17 with the exact prior-week days 8-10 or exact following-week days 22-24. Each placebo retains all three case days and 396 matched strata. The mid-month estimates are smaller than the days 1-3 estimate and have opposite signs (-1.1% and +1.7%); both confidence intervals include the null. Bottom rows: the primary case-crossover re-fit within three sub-periods of the study window.

Sensitivity	Strata	IRR	CI_lo	CI_hi	P
Primary: days 1-3 vs same-weekday days 10-24	396	1.0861	1.0743	1.0979	< 1e-15
Placebo: days 15-17 vs exact prior-week days 8-10	396	0.9893	0.9782	1.0006	0.0639
Placebo: days 15-17 vs exact following-week days 22-24	396	1.0166	0.9999	1.0335	0.0519
Years 2012-2015	132	1.1061	1.0848	1.1279	< 1e-15
Years 2016-2019	132	1.0856	1.0665	1.1050	< 1e-15
Years 2020-2023	132	1.0672	1.0474	1.0874	1.11e-11

Extended Data Table 5. Celebration endpoint counts and category-share models. Panel A reports count IRRs for the eight primary endpoint-event cells. Headline inference uses the cluster-robust standard error with a t critical value and G - 1 degrees of freedom; normal and paired event-cluster bootstrap intervals are sensitivities. Panel B reports adjusted odds ratios comparing each category versus all other calls between event and control windows, using the same treatment, weather covariates, fixed effects and event clusters. These are not percentage-point changes in share.

Panel A: endpoint-specific count models

Event	Window	Endpoint	IRR	95% CI (t, G-1 df; headline)	95% CI (normal sensitivity)	95% CI (bootstrap sensitivity)	P (t, G-1 df)
July 4	Fireworks window	Assault	2.329	[1.92, 2.83]	[1.96, 2.77]	[1.98, 2.80]	1.15e-06
July 4	Fireworks window	Burns	3.335	[2.44, 4.56]	[2.52, 4.41]	[2.46, 4.30]	3.84e-06
July 4	Fireworks window	Cardiac Arrest	1.058	[0.97, 1.16]	[0.98, 1.15]	[0.96, 1.13]	0.202
July 4	Fireworks window	Drug or poisoning	1.341	[1.21, 1.49]	[1.22, 1.47]	[1.23, 1.48]	6.06e-05
Memorial + Labor Day	Holiday day	Assault	1.393	[1.25, 1.55]	[1.26, 1.55]	[1.25, 1.53]	2.16e-06
Memorial + Labor Day	Holiday day	Cardiac Arrest	0.999	[0.93, 1.07]	[0.93, 1.07]	[0.93, 1.07]	0.968
Memorial + Labor Day	Holiday day	Mental Health	0.842	[0.82, 0.87]	[0.82, 0.87]	[0.82, 0.87]	7.54e-11
Memorial + Labor Day	Holiday day	Drug or poisoning	1.164	[1.10, 1.23]	[1.10, 1.23]	[1.11, 1.23]	1.59e-05

Panel B: direct endpoint-share models

Event	Window	Endpoint	Share OR	95% CI (t, G-1 df)	P
July 4	Fireworks window	Burns	2.833	[2.07, 3.87]	1.44e-05
July 4	Fireworks window	Assault	1.892	[1.60, 2.24]	3.93e-06
July 4	Fireworks window	Drug or poisoning	1.126	[1.04, 1.23]	0.01
July 4	Fireworks window	Cardiac arrest	0.899	[0.81, 1.00]	0.0492
Memorial + Labor Day	Holiday day	Assault	1.407	[1.27, 1.55]	3.05e-07
Memorial + Labor Day	Holiday day	Drug or poisoning	1.211	[1.15, 1.27]	3.4e-08
Memorial + Labor Day	Holiday day	Mental health	0.901	[0.87, 0.93]	7.17e-07
Memorial + Labor Day	Holiday day	Cardiac arrest	1.083	[1.01, 1.17]	0.0345

Extended Data Table 6. Memorial and Labor Day mental-health calls by time of day. Dispatcher-coded mental-health calls were lower during business hours (9-17), evening hours (18-23) and overnight (0-8), with the largest reduction during business hours. Confidence intervals and P values use the cluster-robust standard error with a t critical value and 23 degrees of freedom for the 24 holiday occurrences. This timing is compatible with, but not diagnostic of, changes in care routing or dispatch classification; the data do not show a reduction in mental-health need.

Hour band	% change	CI	P	Treated rate/hr	Control rate/hr
Full day (00-23 h)	-15.8%	[-18.4%, -13.1%]	7.54e-11	14.43	17.07
Business hours (09-17 h)	-22.4%	[-25.7%, -19.0%]	1.64e-11	17.94	23.11
Evening (18-23 h)	-10.0%	[-15.1%, -4.6%]	0.00106	17.72	19.61
Overnight (00-08 h)	-7.5%	[-13.2%, -1.4%]	0.0183	8.72	9.33

Extended Data Table 7. Marathon complete-interval estimates by income quartile and cell counts. Stratified 1 km vs >5 km difference-in-differences estimates for the complete call-to-hospital interval among records with completed transport, alongside the incident counts identifying each cell. Income quartiles are defined in Supplementary Methods. Estimates are positive in every quartile but do not follow a monotonic income gradient. The highest-income quartile has the fewest marathon-day incidents in control ZIPs (288) because many high-income ZIPs lie near the route. In the continuous-distance specification of Extended Data Table 8, the highest-income marathon-day change is +143 seconds at 0.5 km (95% CI 52-233) and -65 seconds at 5 km (-172 to 42). Neither specification establishes between-quartile differences.

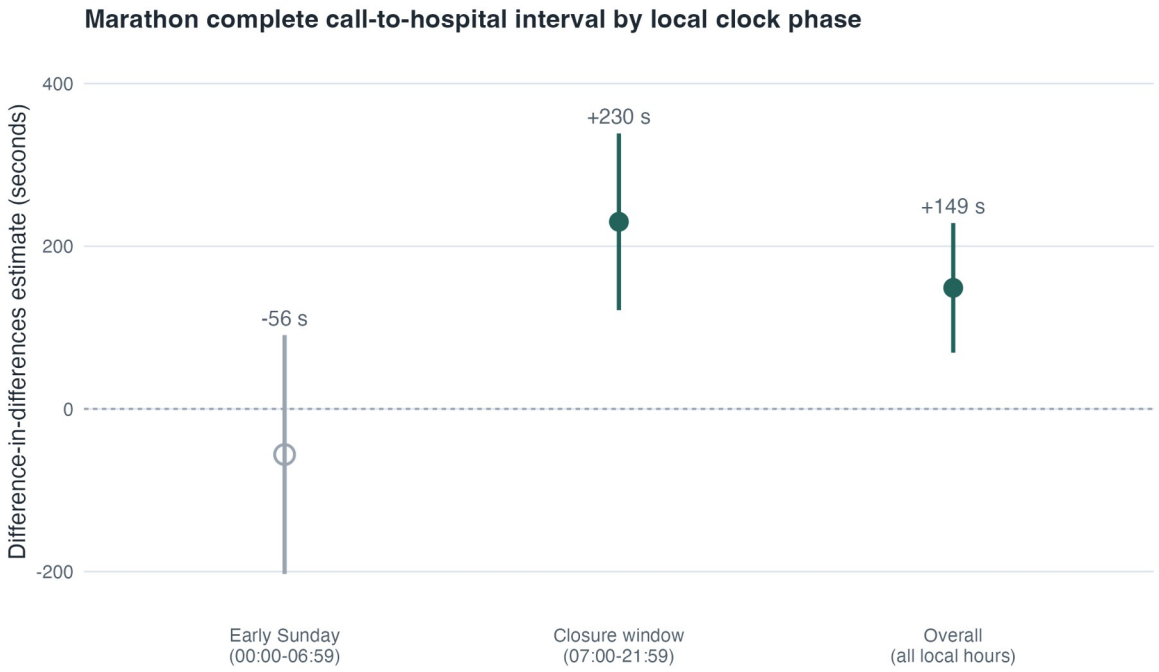
Quartile	Estimate (s; P)	Event-near	Event-control	Control-near	Control-control
Q1	145 (0.00895)	1308	3833	5336	15030
Q2	200 (0.0211)	353	3493	1307	14106
Q3	192 (0.00359)	613	2182	2204	9102
Q4	199 (0.000594)	1184	288	3936	1110

Extended Data Table 8. Continuous distance-by-income-quartile interaction for the complete interval. Marginal marathon-day estimates for the complete call-to-hospital interval among records with completed transport at selected distances from the route, separately by neighborhood income quartile, from a difference-in-differences model interacting the marathon-day indicator with a natural spline of log distance (3 degrees of freedom) and ZIP income quartile. Lowest- and third-quartile estimates are positive at every displayed distance

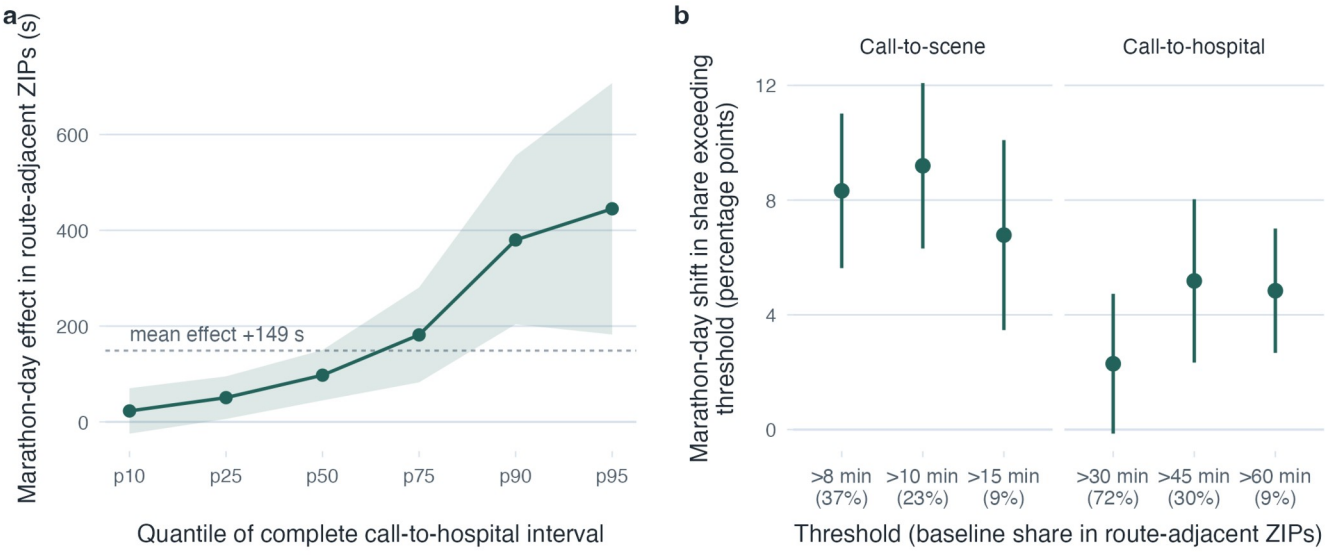
and second-quartile estimates within 1 km; highest-income intervals include zero at every displayed distance. Differential precision does not establish a monotonic gradient or greater burden in one quartile.

Income quartile	Distance (km)	Estimate (s)	95% CI	P	Identification
Q1	0.5	225	[118, 331]	3.41e-05	ESTIMATED
Q1	1.0	194	[58, 331]	0.00528	ESTIMATED
Q1	2.0	161	[27, 294]	0.0184	ESTIMATED
Q1	5.0	109	[47, 171]	0.000521	ESTIMATED
Q2	0.5	287	[111, 462]	0.00134	ESTIMATED
Q2	1.0	298	[98, 499]	0.00354	ESTIMATED
Q2	2.0	265	[84, 445]	0.00403	ESTIMATED
Q2	5.0	142	[62, 223]	0.000542	ESTIMATED
Q3	0.5	255	[86, 425]	0.00313	ESTIMATED
Q3	1.0	234	[44, 424]	0.016	ESTIMATED
Q3	2.0	195	[26, 364]	0.0239	ESTIMATED
Q3	5.0	113	[20, 205]	0.0166	ESTIMATED
Q4	0.5	143	[52, 233]	0.00205	ESTIMATED
Q4	1.0	94	[9, 179]	0.0294	ESTIMATED
Q4	2.0	34	[-47, 115]	0.411	ESTIMATED
Q4	5.0	-65	[-172, 42]	0.235	ESTIMATED

Extended Data Figure 1. Legacy early-Sunday phase comparison for the complete interval. For the completed-transport call-to-hospital outcome, a treated-control difference was not detected before 07:00 and was detected in the modelled 07:00-21:59 window. Because some route segments close before 07:00, the early-Sunday interval is not interpreted as a clean pre-closure placebo; the principal-outcome prior-Saturday comparison is shown in Fig. 1c.



Extended Data Figure 2. The complete marathon interval is concentrated in the tail of the response-time distribution. (a) Unconditional quantile (recentered influence function) difference-in-differences estimates of the marathon-day association with the complete call-to-hospital interval among records with completed transport in route-adjacent ZIPs, at the 10th to 95th percentiles, with the mean estimate (+149 s) shown as a dashed line. Estimates rise from +23 s at the 10th percentile (P = 0.35) to +445 s at the 95th. (b) Linear probability difference-in-differences estimates of the marathon-day shift in the share of incidents exceeding fixed thresholds of call-to-scene response (dispatch + travel; >8, >10, >15 min) and of the complete call-to-hospital interval (>30, >45, >60 min), with route-adjacent control-day baseline shares in parentheses. Shading and vertical bars show 95% confidence intervals; standard errors are clustered by ZIP and date throughout.



Supplementary Tables

Supplementary Table S1. Sensitivity to COVID-19 period exclusion. Early-month drug or poisoning estimate recalculated after excluding 2020-2021. The estimate remains comparable to the full sample (IRR 1.091, 95% CI 1.078-1.103 vs 1.086 in the full sample).

Analysis	IRR	95% CI	P	Case-day strata
Early-month drug or poisoning; 2020-2021 excluded	1.091	[1.078, 1.103]	< 1e-15	330

Supplementary Table S2. Exploratory celebration endpoint x event x window results with Benjamini-Hochberg q-values. All endpoint x event x window tests outside the prespecified primary set, with cluster-robust normal CIs, small-cluster t(df = G - 1) CIs, and Benjamini-Hochberg q-values controlling the false-discovery rate. The leading exploratory cells include New Year's Eve drug or poisoning, trauma and assault.

Event	Window	Endpoint	G	IRR	CI (small-G)	P (small-G)	q (BH-FDR)
Memorial + Labor Day	daytime	Total Calls	24	0.793	[0.77, 0.82]	1.88e-13	1.69e-11

Event	Window	Endpoint	G	IRR	CI (small-G)	P (small-G)	q (BH-FDR)
Memorial + Labor Day	daytime	Trauma	24	0.756	[0.73, 0.79]	7.24e-13	3.26e-11
Memorial + Labor Day	daytime	Mental Health	24	0.750	[0.71, 0.79]	3.95e-11	1.19e-09
Memorial + Labor Day	overnight	Drug or poisoning	24	1.441	[1.34, 1.54]	1.72e-10	3.88e-09
July 4	After midnight	Trauma	12	1.784	[1.63, 1.95]	1.84e-08	3.31e-07
New Year	Early morning	Drug or poisoning	12	5.183	[4.00, 6.71]	2.39e-08	3.58e-07
Memorial + Labor Day	afternoon	Mental Health	24	0.828	[0.79, 0.87]	4.76e-08	6.12e-07
New Year	Core window	Drug or poisoning	12	2.982	[2.42, 3.67]	1.61e-07	1.81e-06
New Year	Late morning	Drug or poisoning	12	2.491	[2.09, 2.97]	1.98e-07	1.98e-06
July 4	Fireworks window	Trauma	12	1.429	[1.33, 1.54]	3.02e-07	2.72e-06
July 4	After midnight	Assault	12	2.888	[2.31, 3.61]	4.6e-07	2.95e-06
July 4	After midnight	Total Calls	12	1.349	[1.27, 1.44]	4.38e-07	2.95e-06
New Year	Early morning	Total Calls	12	2.365	[1.98, 2.83]	4.38e-07	2.95e-06
Memorial + Labor Day	overnight	Assault	24	1.878	[1.56, 2.27]	4.39e-07	2.95e-06
July 4	daytime	Trauma	12	0.756	[0.71, 0.80]	5.37e-07	3.22e-06
Memorial + Labor Day	overnight	Trauma	24	1.221	[1.15, 1.30]	7.3e-07	4.11e-06
July 4	Fireworks window	Total Calls	12	1.191	[1.15, 1.24]	8.11e-07	4.29e-06
New Year	Early morning	Assault	12	7.057	[4.47, 11.15]	1.35e-06	6.75e-06
New Year	Core window	Assault	12	4.047	[2.90, 5.64]	1.59e-06	6.84e-06
New Year	Core window	Total Calls	12	1.569	[1.41, 1.75]	1.54e-06	6.84e-06
New Year	Early morning	Trauma	12	3.408	[2.55, 4.56]	1.51e-06	6.84e-06
Memorial + Labor Day	Holiday day	Total Calls	24	0.932	[0.91, 0.95]	2.28e-06	9.31e-06
July 4	Holiday day	Mental Health	12	0.830	[0.79, 0.87]	2.49e-06	9.33e-06
Memorial + Labor Day	overnight	Total Calls	24	1.102	[1.07, 1.14]	2.43e-06	9.33e-06
Memorial + Labor Day	evening	Trauma	24	1.162	[1.10, 1.22]	4.98e-06	1.79e-05
Memorial + Labor Day	afternoon	Total Calls	24	0.913	[0.88, 0.94]	5.27e-06	1.83e-05
New Year	Core window	Cardiac Arrest	12	1.413	[1.28, 1.55]	7.04e-06	2.3e-05
New Year	countdown	Drug or poisoning	12	1.663	[1.44, 1.92]	7.4e-06	2.3e-05
New Year	Early morning	Cardiac Arrest	12	1.905	[1.59, 2.28]	7.38e-06	2.3e-05
New Year	Core window	Trauma	12	1.905	[1.59, 2.29]	8.78e-06	2.63e-05
July 4	daytime	Total Calls	12	0.803	[0.75, 0.86]	1.51e-05	4.38e-05
New Year	Late morning	Assault	12	3.072	[2.13, 4.43]	3.17e-05	8.93e-05
July 4	evening	Trauma	12	1.246	[1.16, 1.34]	3.29e-05	8.97e-05
July 4	evening	Burns	12	3.244	[2.20, 4.78]	3.4e-05	9e-05

Event	Window	Endpoint	G	IRR	CI (small-G)	P (small-G)	q (BH-FDR)
July 4	evening	Total Calls	12	1.099	[1.06, 1.14]	5.14e-05	0.000132
July 4	evening	Drug or poisoning	12	1.261	[1.16, 1.37]	8.2e-05	0.000205
July 4	After midnight	Burns	12	3.587	[2.24, 5.74]	9.05e-05	0.00022
July 4	daytime	Mental Health	12	0.799	[0.73, 0.87]	0.00011	0.00026
July 4	evening	Assault	12	1.840	[1.44, 2.34]	0.000174	0.000394
July 4	After midnight	Drug or poisoning	12	1.469	[1.26, 1.71]	0.000175	0.000394
July 4	Holiday day	Burns	12	1.822	[1.43, 2.32]	0.000184	0.000404
July 4	Holiday day	Total Calls	12	0.944	[0.92, 0.97]	0.000209	0.000438
New Year	Late morning	Mental Health	12	0.748	[0.67, 0.84]	0.000209	0.000438
Memorial + Labor Day	afternoon	Trauma	24	0.890	[0.84, 0.94]	0.000229	0.000468
July 4	Holiday day	Drug or poisoning	12	1.223	[1.12, 1.33]	0.000285	0.000569
July 4	Holiday day	Assault	12	1.460	[1.24, 1.72]	0.000301	0.000589
Memorial + Labor Day	evening	Total Calls	24	1.052	[1.03, 1.08]	0.000482	0.000923
New Year	countdown	Total Calls	12	1.057	[1.03, 1.09]	0.00114	0.00214
New Year	countdown	Mental Health	12	0.865	[0.80, 0.93]	0.00142	0.00261
Memorial + Labor Day	Holiday day	Trauma	24	0.946	[0.92, 0.98]	0.0016	0.00288
Memorial + Labor Day	evening	Drug or poisoning	24	1.122	[1.05, 1.20]	0.00201	0.00356
Memorial + Labor Day	evening	Assault	24	1.326	[1.12, 1.58]	0.00266	0.00461
July 4	evening	Mental Health	12	0.851	[0.77, 0.94]	0.00386	0.00655
New Year	Early morning	Burns	12	2.943	[1.48, 5.84]	0.00529	0.00881
New Year	countdown	Assault	12	1.691	[1.20, 2.38]	0.00589	0.00964
July 4	Fireworks window	Mental Health	12	0.913	[0.85, 0.98]	0.0121	0.0195
July 4	After midnight	Cardiac Arrest	12	1.214	[1.04, 1.42]	0.0205	0.0323
Memorial + Labor Day	afternoon	Drug or poisoning	24	1.090	[1.01, 1.18]	0.0383	0.0594
Memorial + Labor Day	evening	Mental Health	24	0.932	[0.87, 1.00]	0.04	0.061
New Year	Early morning	Mental Health	12	1.272	[1.01, 1.61]	0.0454	0.067
New Year	Late morning	Burns	12	2.356	[1.02, 5.43]	0.0451	0.067
New Year	Core window	Burns	12	1.409	[0.98, 2.02]	0.0607	0.0874
Memorial + Labor Day	afternoon	Assault	24	1.213	[0.99, 1.49]	0.0612	0.0874
July 4	daytime	Drug or poisoning	12	1.091	[0.99, 1.20]	0.0686	0.0965
New Year	countdown	Cardiac Arrest	12	1.094	[0.99, 1.21]	0.0719	0.0995
July 4	Holiday day	Trauma	12	0.972	[0.94, 1.01]	0.0935	0.127
Memorial + Labor Day	overnight	Burns	24	1.390	[0.88, 2.20]	0.15	0.202
New Year	countdown	Trauma	12	1.058	[0.98, 1.15]	0.156	0.206

Event	Window	Endpoint	G	IRR	CI (small-G)	P (small-G)	q (BH-FDR)
Memorial + Labor Day	afternoon	Cardiac Arrest	24	0.940	[0.85, 1.03]	0.189	0.247
Memorial + Labor Day	daytime	Burns	24	0.811	[0.58, 1.14]	0.218	0.28
Memorial + Labor Day	evening	Cardiac Arrest	24	1.080	[0.92, 1.26]	0.326	0.413
July 4	Holiday day	Cardiac Arrest	12	1.025	[0.96, 1.09]	0.411	0.506
New Year	Late morning	Cardiac Arrest	12	1.083	[0.88, 1.33]	0.416	0.506
Memorial + Labor Day	overnight	Mental Health	24	0.975	[0.92, 1.04]	0.408	0.506
Memorial + Labor Day	overnight	Cardiac Arrest	24	1.039	[0.92, 1.17]	0.507	0.608
July 4	After midnight	Mental Health	12	1.034	[0.92, 1.16]	0.529	0.627
Memorial + Labor Day	afternoon	Burns	24	1.104	[0.79, 1.53]	0.539	0.63
July 4	evening	Cardiac Arrest	12	0.971	[0.87, 1.08]	0.552	0.637
Memorial + Labor Day	daytime	Cardiac Arrest	24	0.971	[0.87, 1.08]	0.583	0.663
Memorial + Labor Day	daytime	Drug or poisoning	24	1.025	[0.93, 1.13]	0.59	0.663
July 4	daytime	Assault	12	0.915	[0.60, 1.39]	0.646	0.711
New Year	countdown	Burns	12	0.880	[0.48, 1.60]	0.648	0.711
Memorial + Labor Day	evening	Burns	24	1.062	[0.80, 1.41]	0.664	0.72
New Year	Late morning	Trauma	12	1.029	[0.89, 1.20]	0.678	0.726
July 4	daytime	Cardiac Arrest	12	1.027	[0.88, 1.19]	0.711	0.744
Memorial + Labor Day	daytime	Assault	24	1.046	[0.81, 1.35]	0.719	0.744
Memorial + Labor Day	Holiday day	Burns	24	1.031	[0.87, 1.22]	0.713	0.744
New Year	Late morning	Total Calls	12	0.993	[0.92, 1.07]	0.846	0.865
New Year	Core window	Mental Health	12	1.008	[0.89, 1.14]	0.884	0.894
July 4	daytime	Burns	12	1.022	[0.64, 1.64]	0.92	0.92

Supplementary Table S3. Seasonal stability of the early-month effect. Monthly case-crossover drug or poisoning IRRs stratified by season. Because January is excluded throughout the early-month design, the stratum labelled winter contains December and February only. The estimate is largest in summer and smallest, but remains slightly above the null, in that two-month winter stratum.

Season	IRR	95% CI	P	Case-day strata	Case days	Control days
Winter	1.027	[1.002, 1.052]	0.0369	72	72	168
Spring	1.090	[1.062, 1.118]	6.08e-11	108	108	252
Summer	1.106	[1.087, 1.125]	< 1e-15	108	108	252
Fall	1.081	[1.056, 1.106]	7.25e-11	108	108	252

Supplementary Table S4. Marathon complete-interval spatial dose-response. Marginal marathon-day estimates for the complete call-to-hospital interval among records with completed transport at selected

distances from the route, from a difference-in-differences model with a natural-cubic-spline interaction in log distance. The estimate attenuates with distance, showing spatial concentration near the route without isolating the operative event mechanism.

Distance_km	Estimate	CI_lo	CI_hi
0.5	213.5	143.4	283.7
1.0	188.6	122.4	254.8
2.0	156.8	102.6	211.0
5.0	102.8	70.1	135.5
10.0	57.1	15.3	99.0

Supplementary Table S5. Marathon incident-level record-structure verification. Average number of records per unique (ZIP, incident timestamp) key in each cell of the corrected marathon difference-in-differences design. Multi-record keys are below 0.03% in every cell. This makes material inflation from repeated timestamp keys unlikely but does not independently validate the source record structure.

Marathon day	Route-adjacent ZIP	N unique timestamp keys	N records	Mean records/key	% multi-record keys
0	0	56427	56433	1.000106	0.0106332
0	1	19303	19304	1.000052	0.0051805
1	0	14534	14537	1.000206	0.0206413
1	1	5719	5722	1.000525	0.0524567

Supplementary Table S6. Marathon robustness, inference and clock-label checks. Panel A reports the corrected local-clock linear specification for the complete call-to-hospital interval, with sensitivity to a log-response specification, 1st/99th-percentile winsorization and exclusion of incidents before 03:00. All records occurring during the repeated local 01:00 hour were retained, but the public field does not permit assignment to the first versus second physical occurrence. Panel B reports the two-way clustered headline estimates and one-way date- and ZIP-clustered checks. The two-way covariance matrices required positive-semidefinite repair, whereas the one-way matrices did not. Panel C compares local-calendar estimates with a diagnostic that assigns the same current incidents and corrected weather using UTC date/hour labels; these are point-estimate differences, without a joint confidence interval.

Panel A: outcome and DST sensitivity

Specification	Unit	Estimate	CI lo	CI hi	P	Note
Primary (linear)	seconds	148.80	69.10	228.50	0.000618	linear complete interval; corrected local calendar
log	percent	5.04	2.14	8.02	0.00119	log complete interval; equivalent 125.2 s at sample mean 2484.6 s
Winsorized (1st-99th percentile)	seconds	143.30	70.35	216.25	0.000343	winsorized at 909.0 and 5629.9 seconds
Exclude incidents before 03:00 (DST)	seconds	173.77	95.46	252.09	6.9e-05	n = 58,365; removes the repeated fall-back hour

Panel B: clustering sensitivity

Outcome	Clustering	Effect (s)	95% CI	P	PSD repair
Call-to-scene	ZIP+date	+109.8	[56.8, 162.8]	0.000176	TRUE
Call-to-scene	date	+109.8	[63.6, 156.0]	2.5e-05	FALSE
Call-to-scene	ZIP	+109.8	[75.7, 143.9]	5.12e-09	FALSE
Complete call-to-hospital	ZIP+date	+148.8	[69.1, 228.5]	0.000618	TRUE
Complete call-to-hospital	date	+148.8	[71.2, 226.4]	0.000453	FALSE
Complete call-to-hospital	ZIP	+148.8	[94.6, 203.0]	3.85e-07	FALSE

Panel C: same-source UTC-label diagnostic

Outcome	UTC-label effect (s)	Local-label effect (s)	Local minus UTC (s)	Inference
All Sunday, call-to-scene	+101.4	+109.8	+8.4	point difference only; joint CI not estimated
All Sunday, complete call-to-hospital	+143.1	+148.8	+5.7	point difference only; joint CI not estimated

Supplementary Table S7. Early-month pulse amplitude by borough and income quartile. Days 1-3 drug or poisoning IRR (vs days 10-24 of the same month) from event-time Poisson models with day-of-week adjustment, month-year fixed effects and month-year-clustered standard errors, fitted citywide and separately by borough and ZIP-level income quartile (January anchors excluded; quartiles defined in Supplementary Methods). The pulse is present in every borough and every income quartile. Amplitudes differ across boroughs (Cochran's Q = 25.2, df = 4, P < 0.001; largest in the Bronx, smallest in Queens) and income quartiles (Q = 16.9, df = 3, P < 0.001), declining from the lowest- to the highest-income quartile.

Scheme	Subgroup	IRR	95% CI	P	Baseline calls/day
Citywide	All	1.088	[1.072, 1.104]	< 1e-15	252.4
Borough	BRONX	1.126	[1.101, 1.151]	< 1e-15	47.8
Borough	BROOKLYN	1.101	[1.078, 1.123]	< 1e-15	65.7
Borough	MANHATTAN	1.081	[1.060, 1.102]	8.28e-15	83.0
Borough	QUEENS	1.042	[1.020, 1.066]	0.000232	46.6
Borough	RICHMOND / STATEN ISLAND	1.092	[1.051, 1.134]	6.14e-06	9.2
Income quartile	Q1	1.117	[1.098, 1.136]	< 1e-15	98.1
Income quartile	Q2	1.079	[1.058, 1.100]	6.11e-14	62.3
Income quartile	Q3	1.065	[1.040, 1.091]	2.59e-07	40.0
Income quartile	Q4	1.063	[1.040, 1.087]	4.49e-08	51.2

Supplementary Table S8. Marathon unconditional quantile difference-in-differences estimates. Recentered influence function regressions of the marathon-day contrast for the complete call-to-hospital interval in route-adjacent ZIPs at the 10th to 95th percentiles, with the mean difference-in-differences estimate for reference. Covariates, fixed effects and ZIP-and-date clustering match the original complete-interval model.

Quantile	Quantile value (s)	Effect (s)	CI lo	CI hi	P
p10	1464	23	-25	70	0.349
p25	1849	51	6	95	0.0302
p50	2345	98	45	150	0.000671
p75	2950	182	83	281	0.000754
p90	3643	380	204	556	0.000102
p95	4192	445	182	707	0.00169
mean		149	69	229	0.000618

Supplementary Table S9. Marathon threshold-exceedance difference-in-differences estimates. Linear probability models for the marathon-day shift in the share of route-adjacent incidents exceeding fixed response-time thresholds. Call-to-scene response (dispatch + travel) uses the principal validity rule (positive and shorter than 3,600 seconds); the complete call-to-hospital interval is estimated in the secondary completed-transport sample. Baseline shares are route-adjacent control-day means.

Outcome	Threshold	Baseline share	Shift (pp)	CI lo	CI hi	P
Call-to-scene response	>8 min	37.0%	+8.3	5.6	11.0	1.93e-07
Call-to-scene response	>10 min	23.2%	+9.2	6.3	12.1	9.65e-08
Call-to-scene response	>15 min	8.8%	+6.8	3.5	10.1	0.000205
Complete call-to-hospital interval	>30 min	72.1%	+2.3	-0.1	4.7	0.071
Complete call-to-hospital interval	>45 min	29.9%	+5.2	2.3	8.0	0.000819
Complete call-to-hospital interval	>60 min	9.0%	+4.8	2.7	7.0	6.29e-05

Supplementary Table S10. Social Security staggered-Wednesday mechanism check. Drug/poisoning and comparison-endpoint IRRs for mid-month payment Wednesdays (2nd/3rd/4th Wednesday, when post-1997 Social Security retirement and disability benefits are disbursed) versus flanking Tuesdays and Thursdays within days 8-28, with month-year fixed effects and month-year-clustered standard errors. Because every 2nd/3rd/4th Wednesday is a payment day, no unexposed mid-month Wednesday exists; assault and mental health benchmark generic day-of-week patterns. None showed a comparable mid-month Wednesday association. This result does not rule out transfer-related mechanisms clustered at the month boundary.

Contrast	Endpoint	IRR	95% CI	P
2nd-4th Wednesdays pooled	drug or poisoning	0.990	[0.982, 0.999]	0.0261
2nd-4th Wednesdays pooled	assault	1.002	[0.977, 1.029]	0.854
2nd-4th Wednesdays pooled	mental health	1.004	[0.996, 1.011]	0.348
2nd-4th Wednesdays pooled	cardiac arrest	1.011	[0.996, 1.026]	0.163
2nd-4th Wednesdays pooled	unconscious	0.999	[0.987, 1.010]	0.819
2nd Wednesday	drug or poisoning	0.986	[0.972, 1.001]	0.0714
3rd Wednesday	drug or poisoning	0.996	[0.981, 1.011]	0.61
4th Wednesday	drug or poisoning	0.989	[0.974, 1.005]	0.164
Wednesdays pooled, January excluded	drug or poisoning	0.992	[0.983, 1.000]	0.0629

Supplementary Table S11. United Nations General Assembly high-level week response-time difference-in-differences. Marathon-style DiD applied to UN General Debate weeks (first five weekdays from the official opening date, 2012-2023) versus the same weekdays 2 and 4 weeks before and after. Treated ZIPs were within

1.5 km of UN headquarters; ZIPs 1.5-3 km away were excluded and the remaining Manhattan ZIPs served as controls. The primary outcome is the valid-response-weighted ZIP-hour mean call-to-scene time among positive intervals below 7,200 seconds. A sensitivity row applies the 3,600-second ceiling used for the marathon principal outcome. Models use ZIP and event-year fixed effects, local hour and day-of-week adjustment, and ZIP-and-date-clustered standard errors. The delay is larger during 07:00-20:00 and loads mainly on assignment-to-scene travel rather than dispatch. The estimate for the virtual 2020 General Debate is null.

Model	Outcome	Effect (s)	95% CI	P value
Primary (all hours)	Call-to-scene response	+113	[52, 174]	.001
Local hours 07-20	Call-to-scene response	+150	[75, 225]	5e-4
Placebo: 2020 virtual GD	Call-to-scene response	-28	[-88, 32]	.372
Sensitivity: 3,600 s ceiling (all hours)	Call-to-scene response	+74	[35, 113]	1e-3
Component (common sample)	Dispatch	+10	[-3, 22]	.133
Component (common sample)	Assignment-to-scene travel	+93	[36, 149]	.003
Component (outcome-valid sample)	Scene	+20	[-8, 49]	.177
Component (outcome-valid sample)	Transport	+39	[18, 61]	.001

Supplementary Table S12. Off-diagonal pattern contrasts. Panel A reports direct estimates of the cells the margin comparison might otherwise assume. Marathon-to-counts uses a Poisson difference-in-differences model on ZIP-day incident counts over the same treated/control ZIPs and Sundays as the principal access design, with and without event-standby records. Early-month-to-response compares days 1-3 with days 10-24; celebration-to-response compares each holiday with matched days +/-7 and 14 days. Panel B removes the entire local 01:00 hour from every candidate Sunday, leaving 23 physical hours on event and control dates without assigning public records to a first versus second fold. Because the count difference-in-differences interaction already compares route-adjacent and control ZIPs on the same dates, any extra hour is common to both spatial groups; the common-window sensitivity nevertheless tests the concern directly.

Cell	Effect	95% CI	P
Marathon -> incident counts (all records)	IRR 1.151	[1.052, 1.258]	0.00211
Marathon -> incident counts (excl. standby)	IRR 1.152	[1.054, 1.260]	0.00184
Marathon -> cardiac-arrest counts	IRR 1.247	[1.012, 1.537]	0.0385
Early month -> call-to-scene	+5.3 s	[-5.0, 15.6]	0.311
July 4 -> call-to-scene	-67.5 s	[-93.8, -41.3]	0.000379
Memorial/Labor Day -> call-to-scene	-89.4 s	[-113.5, -65.4]	1.58e-05

Panel B: daylight-saving-time-balanced count sensitivity

Window	Outcome	Estimate and inference
All physical hours	All incident records	IRR 1.151 [1.052, 1.258]; P = 0.00211
All physical hours	Excluding standby	IRR 1.152 [1.054, 1.260]; P = 0.00184
All physical hours	Cardiac arrest	IRR 1.247 [1.012, 1.537]; P = 0.0385

Window	Outcome	Estimate and inference
Common 23 hours (local 01:00 excluded from every date)	All incident records	IRR 1.163 [1.059, 1.278]; P = 0.00168
Common 23 hours (local 01:00 excluded from every date)	Excluding standby	IRR 1.165 [1.060, 1.280]; P = 0.00148
Common 23 hours (local 01:00 excluded from every date)	Cardiac arrest	IRR 1.294 [1.035, 1.618]; P = 0.0236

Supplementary Table S13. Marathon complete-interval timing relative to the early-month window and baseline EMS demand. Every held NYC Marathon falls within days 1-7 of November; four of ten fall in the days 1-3 drug or poisoning peak. Top rows report the route-adjacent complete call-to-hospital estimate separately for marathons in days 1-3 and days 4-7; the confidence intervals overlap and the peak split contains only four events. Bottom rows stratify the complete-interval estimates by neighbourhood baseline EMS demand. Mid- and high-demand strata are precise whereas the low-demand estimate is imprecise, but the similar mid- and high-demand point estimates do not establish a dose gradient. Baseline demand and deprivation also coincide in the route-adjacent set, so this is not an independent interaction.

The preserved estimates use normal-reference confidence intervals and clustered t-test P values.

Stratum	Effect (s)	95% CI (normal)	P (t test)
All marathons	+149	[69, 229]	0.000618
Marathon in days 1-3 (peak)	+110	[26, 194]	0.0184
Marathon in days 4-7 (tail)	+179	[91, 266]	0.000387
Low baseline demand (4.4 calls/day)	+165	[-5, 335]	0.0651
Mid baseline demand (12.6 calls/day)	+143	[44, 243]	0.00743
High baseline demand (30.9 calls/day)	+153	[70, 236]	0.000848

Supplementary Table S14. Marathon interval validity and sample selection. The secondary complete call-to-hospital sample drops 31.6% of candidate incidents, 99.2% of them incidents without a completed transport. Complete-interval validity was 2.7 percentage points lower differentially in route-adjacent ZIPs on marathon days. Call-to-scene time is available for 98.7% of candidates, but its validity was 3.7 percentage points lower differentially near the route on marathon days. This imbalance motivated the record-level invalid-interval audit and deterministic selection stress tests in Supplementary Table S19.

Panel A: sample retention

Quantity	Value
Candidate incidents	95,996
Valid call-to-scene	94,768 (98.7%)
Valid complete interval	65,639 (68.4%)
Complete-interval exclusions	30,357 (31.6%)
No transport among complete-interval exclusions	99.2%

Panel B: differential interval validity

Outcome	Window	Validity DiD (pp)	95% CI	P
Valid call-to-scene	All hours	-3.7	[-6.0, -1.3]	0.00336
Valid complete interval	All hours	-2.7	[-4.8, -0.6]	0.0142
Valid complete interval	Hours 09-16	-4.3	[-7.4, -1.1]	0.0106
No transport	All hours	+2.4	[0.5, 4.3]	0.0155

Supplementary Table S15. Payment-date displacement probes SSI timing relative to the calendar boundary. SSI and legacy Social Security payments scheduled on weekends or federal holidays are advanced to the preceding business day. The persistence of the days 1-3 contrast in shifted SSI months is inconsistent with a pure payment-date explanation, while the pre-boundary SSI interval supports an SSI-aligned component. We did not detect retiming after legacy Social Security advances. Together these results support a mixed SSI-aligned and calendar-boundary pattern without ruling out other transfer-related mechanisms at the month boundary.

Test	IRR (95% CI)	P
SSI: days 1-3, shifted/unshifted IRR ratio	1.005 [0.974, 1.037]	0.754
SSI: advanced payment through day before 1st	1.070 [1.037, 1.104]	<0.001
Legacy Social Security: shifted/unshifted waveform ratio	0.992 [0.959, 1.025]	0.633

Supplementary Table S16. Outcome-independent scheduled street-event generalization test. NYC historical permit records defined the 2024-2025 event universe without reference to EMS outcomes. Multi-precinct full-street parades and races were associated with longer call-to-scene time through dispatch and assignment-to-scene travel; scene time, local full closures and less-restrictive permits were null. Formal class contrasts differed from the local and less-restrictive classes, and the joint test of the shifted windows did not reject the null (Panel A, last row). Panel B shows the parent estimate by year and its range after each event-date omission. Panel C applies the parent fixed-effects structure to call-to-scene validity. The design does not separate physical closure from crowds, staging, policing or other event operations.

Panel A: parent estimates and class contrasts

Contrast	Effect (s)	95% CI	P
Multi-precinct class: call-to-scene	+37.4	[15.8, 59.1]	0.00115
Multi-precinct class: dispatch	+13.0	[3.8, 22.2]	0.00718
Multi-precinct class: travel	+24.4	[6.6, 42.3]	0.00904
Multi-precinct class: scene	-11.0	[-39.8, 17.9]	0.459
Immediately preceding window	+1.8	[-12.9, 16.5]	0.811
Network minus local full closure	+38.2	[11.1, 65.3]	0.00573
Network minus less-restrictive	+37.4	[16.1, 58.7]	0.000574
Local full minus less-restrictive	-0.8	[-16.0, 14.4]	0.917
Joint test of 5 shifted windows (chi-squared 5.25, 5 df)			0.386

Panel B: year and influence stability

Check	Effect (s)	95% CI or range	P
Year 2024	+42.8	[14.7, 70.9]	0.00382
Year 2025	+30.8	[-4.0, 65.7]	0.0871
Leave-one-date range (187 omissions)		35.4 to 40.4 s	

Panel C: parent-specification validity model

Outcome	Validity difference (pp)	95% CI	P
Valid call-to-scene fraction	-0.54	[-1.39, 0.31]	0.217

Supplementary Table S17. Exploratory permitted-event scale analysis. This extension was frozen after the parent permit-census result but before scale-category outcomes were estimated. It restricts the target class to exact parade and athletic-race records classified as full street closures by the parent scheduled-event specification and compares exposure spanning one, two, or at least three precincts under the parent fixed-effects and weighting specification. Panel A reports category estimates; Panel B reports covariance-based contrasts and stratified checks; Panel C reports the three-or-more-precinct component estimates; Panel D reports the joint shifted-window test and leave-one-date range; Panel E reports occurrence counts and analytic coverage by scale and event type after overlap and precinct exclusions. Point estimates are ordered, but the pooled multi-precinct minus single-precinct confidence interval includes zero, so the analysis is suggestive of scale rather than evidence of a dose response.

Panel A: scale-category estimates

Scale	Effect (s)	95% CI	P
1 precinct	+7.3	[-17.4, 32.0]	0.564
2 precincts	+25.1	[-7.1, 57.3]	0.131
3+ precincts	+42.6	[12.0, 73.1]	0.00786

Panel B: covariance-based contrasts

Scope	Contrast	Effect (s)	95% CI	P
Combined	2 minus 1 precinct	+17.8	[-21.8, 57.3]	0.381
Combined	3+ minus 2 precincts	+17.5	[-29.3, 64.3]	0.466
Combined	3+ minus 1 precinct	+35.3	[-2.0, 72.6]	0.0676
Combined pooled	Multi-precinct minus 1 precinct	+30.2	[-1.5, 61.9]	0.0657
Year 2024	3+ minus 1 precinct	+40.2	[-3.4, 83.8]	0.0747
Year 2025	3+ minus 1 precinct	+26.1	[-32.0, 84.3]	0.381
Parades	3+ minus 1 precinct	+27.2	[-13.3, 67.7]	0.192
Athletic races	3+ minus 1 precinct	+84.4	[4.0, 164.8]	0.0431

Panel C: 3+ precinct components

Component	Effect (s)	95% CI	P
Dispatch	+19.1	[6.5, 31.8]	0.00406
Assignment-to-scene travel	+23.5	[-0.6, 47.5]	0.0594
Scene	-15.1	[-49.0, 18.9]	0.388

Panel D: timing and influence checks

Check	Result
Joint test of 15 shifted-window coefficients	chi-square(15) = 13.1; P = 0.597
Leave-one-event-date multi-minus-one estimate range	+25.7 to +34.0 s across 278 omissions

Panel E: event universe and analytic coverage

Event class / scale	Occurrences (dates)	Analytic precinct-hours	Calls (valid call-to-scene %)
Athletic race/tour / 1	60 (48)	117	349 (95.4%)
Parade / 1	490 (228)	766	2502 (93.4%)
Athletic race/tour / 2	24 (23)	104	267 (93.6%)
Parade / 2	191 (125)	748	2149 (91.4%)
Athletic race/tour / 3+	31 (28)	673	1842 (87.1%)
Parade / 3+	158 (111)	2010	6002 (89.9%)

Supplementary Table S18. Dated 2024-2025 out-of-period directional validation of endpoint-count patterns and response-time contrasts. This analysis asks whether calendar patterns recur over time; it is separate from the permit-census generalization test of street-event access. Before extracting endpoint-specific 2025 counts and applying the 2024-2025 response-time extract, outcomes, windows and decision rules were fixed in a dated protocol. Because the 2024 endpoint counts had appeared in an exploratory update, the exercise was only partly prospective; 2025 endpoint counts were untouched. Panel A compares the 2012-2023 and pooled 2024-2025 IRRs from identical unadjusted matched models. Seven nonzero directions entered the locked decision; the other cells are comparison cells. Panel B reports the seven directional cells separately in 2024 and 2025. Panel C reports total-call-weighted call-to-scene contrasts with the locked normal-reference interval and, as plotted in Fig. 7b, a t-reference interval with one fewer degree of freedom than event clusters. The 2012-2023 Memorial and Labor Day contrast clusters by calendar year, so the two holidays in a year share a cluster (12 clusters for 24 occurrences), whereas in 2024-2025 each event is its own cluster. The validation period contains two July 4 and four pooled Memorial/Labor occurrences, so its holiday intervals are descriptive. Panel D reports the three locked decisions: the endpoint-count pattern met its criterion, whereas the calendar-event response-time classifier did not. Because the prespecified rule required both components, the combined signature framework was classified as not supported.

Panel A: complete 15-cell endpoint ledger

Schedule	Endpoint	2012-2023 IRR	2024-2025 IRR	Locked direction	Direction matched
Early month	Assault	1.005	1.021	Specificity comparison	Not locked
Early month	Burns	0.995	1.009	Specificity comparison	Not locked
Early month	Cardiac arrest	1.045	1.036	Specificity comparison	Not locked
Early month	Mental health	1.000	1.000	Specificity comparison	Not locked
Early month	Drug or poisoning	1.086	1.090	Positive	Yes
July 4	Assault	2.373	1.786	Positive	Yes

Schedule	Endpoint	2012-2023 IRR	2024-2025 IRR	Locked direction	Direction matched
July 4	Burns	3.391	1.905	Positive	Yes
July 4	Cardiac arrest	1.071	1.067	Specificity comparison	Not locked
July 4	Mental health	0.916	0.873	Specificity comparison	Not locked
July 4	Drug or poisoning	1.350	1.120	Positive	Yes
Memorial + Labor Day	Assault	1.440	1.119	Positive	Yes
Memorial + Labor Day	Burns	1.062	1.357	Specificity comparison	Not locked
Memorial + Labor Day	Cardiac arrest	1.006	1.003	Specificity comparison	Not locked
Memorial + Labor Day	Mental health	0.845	0.838	Negative	Yes
Memorial + Labor Day	Drug or poisoning	1.181	1.129	Positive	Yes

Panel B: locked directional cells by validation year

Year	Schedule	Endpoint	IRR	95% CI
2024	Early month	Drug or poisoning	1.084	[1.041, 1.128]
2024	July 4	Burns	1.800	Descriptive
2024	July 4	Assault	1.880	Descriptive
2024	July 4	Drug or poisoning	1.272	Descriptive
2024	Memorial + Labor Day	Assault	1.109	Descriptive
2024	Memorial + Labor Day	Drug or poisoning	1.060	Descriptive
2024	Memorial + Labor Day	Mental health	0.830	Descriptive
2025	Early month	Drug or poisoning	1.096	[1.054, 1.140]
2025	July 4	Burns	2.000	Descriptive
2025	July 4	Assault	1.694	Descriptive
2025	July 4	Drug or poisoning	1.002	Descriptive
2025	Memorial + Labor Day	Assault	1.132	Descriptive
2025	Memorial + Labor Day	Drug or poisoning	1.203	Descriptive
2025	Memorial + Labor Day	Mental health	0.846	Descriptive

Panel C: call-to-scene contrasts

Period	Schedule	Effect (s)	95% CI, locked normal reference	95% CI, t reference (Fig. 6b)	Matched sets / event clusters
2012-2023	Early month	+5.3	[-5.0, 15.6]	[-5.1, 15.7]	132
2012-2023	July 4	-67.5	[-93.8, -41.3]	[-97.0, -38.0]	12
2012-2023	Memorial + Labor Day	-89.4	[-113.5, -65.4]	[-116.5, -62.4]	12
2024-2025	Early month	-5.7	[-32.6, 21.1]	[-33.1, 21.6]	66
2024-2025	July 4	+61.7	[37.2, 86.2]	[-97.1, 220.5]	2

Period	Schedule	Effect (s)	95% CI, locked normal reference	95% CI, t reference (Fig. 6b)	Matched sets / event clusters
2024-2025	Memorial + Labor Day	-202.4	[-289.6, -115.2]	[-343.9, -60.8]	4

Panel D: locked decisions

Component and decision	Result
Endpoint-count pattern — Supported	7 of 7 locked directions matched
Locked calendar-event response-time classifier — Not supported	2 of 3 calendar-event directions matched; July 4 reversed
Combined signature framework — Not supported	Requires both components; response-time component not supported

Supplementary Table S19. Marathon call-to-scene invalid-interval audit, selection stress tests and operational-burden translation. Panel A classifies every incident outside the principal 0-3,600-second validity rule by design cell and observed reason. Panel B refits the principal difference-in-differences model after raising the upper ceiling and under deterministic assignments. In the adverse grid, all 316 invalid route-adjacent marathon-day intervals receive the stated common value; invalid records in the other cells receive their observed-valid cell median. The displayed confidence intervals condition on each imposed assignment and are not multiple-imputation intervals. Panel C translates adjusted threshold shifts by multiplying each percentage-point estimate by the 5,406 valid route-adjacent marathon-day incidents across 10 held events. These are descriptive operational counts, not separately fitted incidence estimates.

Panel A: record-level validity audit

Cell	Candidates	Valid, n (%)	Invalid	Observed invalid values
Control Sunday, >5 km	56,433	55,947 (99.1%)	486	351 non-positive; 135 >=3,600 s; 0 non-finite
Control Sunday, <1 km	19,304	19,105 (99.0%)	199	158 non-positive; 41 >=3,600 s; 0 non-finite
Marathon Sunday, >5 km	14,537	14,310 (98.4%)	227	128 non-positive; 95 >=3,600 s; 4 non-finite
Marathon Sunday, <1 km	5,722	5,406 (94.5%)	316	176 non-positive; 140 >=3,600 s; 0 non-finite

Panel B: principal estimate and deterministic sensitivity analyses

Specification	Effect (s)	95% CI	Analytic n
Principal 0-3,600 s rule	+110	[57, 163]	94,768
Positive intervals <4,500 s	+120	[58, 182]	94,916
Positive intervals <7,200 s	+159	[54, 263]	95,074
Positive intervals <10,800 s	+202	[48, 356]	95,147
High positives retained; non-positive values set to cell median	+153	[53, 254]	95,995
High positives retained; event-near non-positive values set to 1 s	+140	[40, 239]	95,995
All 316 invalid event-near intervals set to 1 s	+74	[40, 107]	95,995

Panel C: operational-burden translation

Threshold	Control-day base	Adjusted shift	Relative increase	Descriptive excess across 10 events	Per event
Call-to-scene >8 min	37.0%	+8.3 pp	22.5%	450	45.0
Call-to-scene >10 min	23.2%	+9.2 pp	39.6%	497	49.7
Call-to-scene >15 min	8.8%	+6.8 pp	77.1%	367	36.7

Supplementary Table S20. Timestamp provenance and daylight-saving-time audit. The public incident timestamp is a floating civil time, not a UTC-labelled instant. The frozen pipeline localizes it to America/New_York before conversion to UTC. The repeated fall-back hour is irreducibly ambiguous in the public field; the parser's explicit policy and the affected-row count are reported rather than treating both folds as observed.

Check	Verified evidence
dataset identity	NYC Open Data metadata. 76xm-jjuj; INCIDENT_DATETIME is a floating timestamp [PASS]
frozen raw identity	EMSNYC replication contract. EMS_Incident_Dispatch_Data_20240614.csv; 6260508337 bytes; SHA-256 f70928cbacd8b33b0bc066e8af392402875cfd0878fa8677d83d54ba0e47a271 [PASS]
study period rows	frozen raw scan. 17490489 rows dated 2012-2023 [PASS]
parser semantics	EMSNYC 01a_data_extraction.R lines 32-59. America/New_York wall time localized before conversion to UTC [PASS]
raw to derived example	frozen raw row CAD_INCIDENT_ID 110010790 and incident derivative. 01/01/2011 02:19:47 AM becomes 2011-01-01T07:19:47Z under EST [PASS]
fall back ambiguity	frozen raw scan. 6128 of 17490489 study-period rows occur in repeated fall-back 01:00; parser uses earliest fold and flags them [DISCLOSED]
primary window overlap	calendar and principal marathon specifications. July 4 Memorial/Labor and marathon active/prior-Saturday windows do not overlap repeated 01:00 [PASS]
weather link	calendar primary rerun. weather joined by ZIP and physical UTC hour [PASS]

Supplementary Table S21. Endpoint constructs and response intervals. Chief-complaint categories use the final dispatcher code and are mutually exclusive. They describe information available to dispatch rather than clinical diagnoses.

Construct	Data field or rule	Definition and limitation
Drug or poisoning	final DRUG*/VENOM dispatcher code	aggregates opioid, polysubstance, alcohol-intoxication and envenomation presentations; not opioid-specific; VENOM is excluded in sensitivity analysis
Assault	dispatcher assault codes	dispatcher-coded, not police-confirmed; misses incidents not presented to 911
Burns	dispatcher burn codes	misses cases for which injured people do not use 911
Mental health	dispatcher mental-health/EDP codes	shares boundaries with intoxication and trauma codes
Cardiac arrest	final cardiac-arrest code	comparison endpoint and possible alternative code for an unresponsive patient with unknown cause
Unconscious	final unconscious/unresponsive codes	possible alternative code when no substance is identified
Seizures	dispatcher seizure codes	includes withdrawal- and intoxication-related presentations
Trauma	dispatcher trauma codes	broad injury category reported for New Year's Eve timing only

Construct	Data field or rule	Definition and limitation
All calls	every dispatcher-coded incident in the panel	all-cause benchmark and calendar response-mean weight; not an endpoint in its own right
Dispatch interval	DISPATCH_RESPONSE_SECONDS_QY	call processing to unit assignment
Travel interval	INCIDENT_TRAVEL_TM_SECONDS_QY	unit assignment to first unit on scene
Call-to-scene interval	dispatch plus travel	principal access outcome; does not require completed transport
Scene interval	scene_time_seconds	on-scene to depart-scene
Transport interval	transport_time_seconds	depart-scene to hospital arrival
Complete call-to-hospital interval	sum of dispatch, travel, scene and transport	secondary outcome conditional on completed transport
Incident record	one computer-aided dispatch row	response intervals describe the incident's first responding unit

Supplementary Table S22. Record accounting, source provenance and analysis ledger. Panel A reconciles the frozen raw study-period audit with the two derived analysis branches; the latter are not sequential filters of one another. Panel B identifies the separate 2024-2025 extracts used for calendar validation and scheduled-event generalization. Panel C records the distinct question, specification timing and validation strategy for each analysis. Panel D compares the primary calendar estimates with two reruns of the same estimation code on city-hour counts drawn on 11 November 2026 from the public dispatch data for the same 170 ZIPs. The panel, like its source extraction, counts only incidents with a valid response interval. The first rerun applies the same rule and reproduces the published estimates to within 1.0 percentage point; the second counts all incidents. Counting all incidents adds 3.2% of calls overall: 7.0% of mental-health, 4.7% of assault, 3.4% of drug or poisoning, 3.1% of burn and 0.8% of cardiac-arrest calls. No primary estimate changes direction or whether its interval excludes 1. Among the secondary event-window cells, only July 4 daytime burns changes direction, and its interval includes 1 in every version. Supporting analyses built on the panel were not rerun.

Panel A: primary-period record accounting

Stage and exact rows	Role
Frozen raw snapshot (all dates) (25,984,643 rows)	Source file before study-period restriction
Raw rows dated 2012-2023 (17,490,489 rows)	Timestamp-provenance audit universe
Primary 170-ZIP NLDAS-linked calendar panel (16,514,279 rows)	Primary calendar count and share analyses
Incident-level supporting-analysis branch (16,892,512 rows)	Incident-level clock, acuity and disposition checks

Panel B: separate 2024-2025 source identities

Source	Frozen identity
EMS precinct-hour response extract	Dataset 76xm-jjuj; frozen 2026-07-13 06:00:24 UTC; 1,099,041 rows; 52,683,812 bytes; MD5 51993fcef952b39648cada9f25e2f223

Source	Frozen identity
EMS precinct-hour count extract	Dataset 76xm-jjuj; frozen 2026-07-13 06:02:13 UTC; 1,121,274 rows; 31,335,919 bytes; MD5 f0a6dd0248199772e3bf73a355a39507
EMS city-hour endpoint extract	Dataset 76xm-jjuj; frozen 2026-07-18 06:24:20 UTC; 17,542 rows; 507,768 bytes; MD5 abdb33940c8f7c5972e59df9a204ae1a
NYC permitted-event exact-distinct extract	Dataset bkfu-528j; frozen 2026-07-13 06:10:38 UTC; 21,748 rows; 5,702,704 bytes; MD5 c9863ae8992c31f6d6ab643b5f3cb063

Analysis	Primary exposure and outcome	Specification status	Validation or generalization
Early month	days 1-3; drug or poisoning calls and share	primary calendar contrast; not preregistered	dated 2024-2025 directional validation
Celebrations	July 4 and Memorial/Labor windows; endpoint counts and shares	primary set fixed before exploratory endpoint-window sweep	dated 2024-2025 directional validation
Marathon	route proximity by event date; call-to-scene	principal outcome designated after the complete interval was analysed	UNGA convergence and scheduled-event generalization
Marathon phase and event-level inference	active event window; call-to-scene	dated design frozen before fitting	matched-date permutation and leave-one-event checks
Permit census	permit class; call-to-scene	outcome-independent scheduled-event generalization test	estimates by year and leave-one-date checks
Permit scale extension	precinct span; call-to-scene	exploratory after parent result	no independent validation sample
Calendar validation	locked endpoint directions and calendar response contrasts	dated protocol; 2024 endpoint counts had been viewed, 2025 had not	endpoint pattern supported; calendar response classifier not supported
TLC transport companion	permit-linked trip duration and EMS correlation	exploratory; new design fixed after parent EMS outcomes	selected-trip corroboration; year and shifted-date checks (S26)
Marathon course reconstruction	route-adjacent exposure	course rebuilt along documented streets after review; reran all marathon analyses	length gate; reproduced old classification exactly before replacing it
Fire-company responses	marathon-day fire response near the route	design fixed 19 October 2026 before fitting	2012 and 2020 placebo Sundays (S27, S29)
Clock-change placebo	fall-back Sundays without closures	design fixed before running	reproduction of the 2019 race effect (S29)
Foot traffic (Advan)	presence on early-month days and holidays; adjusted call estimates	design fixed before fitting; licensed source	independent refits (S38)
Temporal calibration	2024-2025 against 2012-2023 log-IRRs	specified after the directional validation was known	descriptive (S39)

Analysis	Primary exposure and outcome	Specification status	Validation or generalization
Labor Day parade check	holiday racial gradient without parade ZIPs	ZIPs fixed from geometry before fitting	reproduction of the published slope (S40)

Panel D: primary calendar estimates with and without the valid-interval rule (% change). Intervals in this panel are normal-reference; the published headline intervals use cluster-robust t references (Extended Data Tables 3 and 5)

Analysis	Cell	Published (valid interval only)	Rerun, valid interval only	Rerun, all incidents
Early month (days 1-3)	Assault	+0.6% (-2.3 to +3.5)	+0.6% (-2.3 to +3.5)	+0.8% (-2.0 to +3.7)
Early month (days 1-3)	Cardiac arrest	+4.5% (+2.4 to +6.6)	+4.4% (+2.3 to +6.5)	+4.3% (+2.2 to +6.4)
Early month (days 1-3)	Drug or poisoning	+8.6% (+7.4 to +9.8)	+8.6% (+7.4 to +9.8)	+8.7% (+7.5 to +9.9)
Early month (days 1-3)	Mental health	+0.1% (-0.7 to +0.8)	+0.0% (-0.8 to +0.8)	+0.1% (-0.7 to +0.8)
Early month (days 1-3)	Seizures	+2.1% (+0.7 to +3.5)	+2.1% (+0.7 to +3.5)	+2.1% (+0.7 to +3.5)
Early month (days 1-3)	Unconscious	+6.2% (+4.7 to +7.7)	+6.1% (+4.7 to +7.7)	+6.3% (+4.8 to +7.8)
Early month by season	Drug or poisoning, fall	+8.1% (+5.6 to +10.6)	+8.1% (+5.6 to +10.6)	+8.5% (+6.0 to +11.0)
Early month by season	Drug or poisoning, spring	+9.0% (+6.2 to +11.8)	+8.9% (+6.1 to +11.8)	+8.9% (+6.2 to +11.7)
Early month by season	Drug or poisoning, summer	+10.6% (+8.7 to +12.5)	+10.6% (+8.7 to +12.5)	+10.7% (+8.7 to +12.6)
Early month by season	Drug or poisoning, winter	+2.7% (+0.2 to +5.2)	+2.6% (+0.1 to +5.1)	+2.6% (+0.1 to +5.2)
July 4 fireworks window	Assault	+132.9% (+95.8 to +176.9)	+132.6% (+95.5 to +176.7)	+129.7% (+93.7 to +172.5)
July 4 fireworks window	Burns	+233.5% (+152.2 to +340.9)	+233.5% (+152.2 to +340.9)	+230.1% (+150.6 to +334.8)
July 4 fireworks window	Cardiac arrest	+5.8% (-2.5 to +14.7)	+5.5% (-2.6 to +14.2)	+6.4% (-1.5 to +15.0)
July 4 fireworks window	Drug or poisoning	+34.1% (+22.3 to +47.0)	+34.3% (+22.2 to +47.7)	+35.6% (+23.6 to +48.8)
Memorial and Labor Day (full day)	Assault	+39.3% (+25.6 to +54.5)	+40.3% (+26.5 to +55.6)	+39.2% (+25.1 to +54.9)
Memorial and Labor Day (full day)	Cardiac arrest	-0.1% (-6.8 to +6.9)	-0.1% (-6.6 to +6.9)	-0.0% (-6.6 to +7.0)
Memorial and Labor Day (full day)	Drug or poisoning	+16.4% (+10.2 to +22.9)	+16.4% (+10.3 to +22.9)	+16.4% (+10.1 to +23.1)
Memorial and Labor Day (full day)	Mental health	-15.8% (-18.3 to -13.3)	-15.7% (-18.2 to -13.2)	-18.3% (-20.8 to -15.8)

Supplementary Table S23. Event-level access evidence, bounded case-mix sensitivities and validity diagnostics. Panel A reports every active-window marathon estimate. Panel B distinguishes the equally weighted annual summary from the incident-weighted model and the two bounded case-mix sensitivities, and gives the incident-weighted prior-Saturday window (21:00-23:59) and the active-minus-prior-Saturday difference shown in Fig. 1c. Panel C reports every in-person United Nations General Assembly estimate and

labels the pooled and equal-year summaries separately. Panel D reports validity diagnostics, including the equal-weight annual active-window marathon validity contrast, and the scheduled-event estimate by year and after each event-date omission. Panel D also applies the Supplementary Table S19 selection stress tests to the equal-weight active-window estimate. In the adverse row, all 278 invalid route-adjacent marathon-day active-window intervals receive one second, and invalid records in the other cells receive their observed-valid cell median. The ceiling row raises the upper validity limit without assignment. These intervals condition on the imposed values and are not multiple-imputation intervals.

Panel A: annual marathon active-window estimates

Event year	Active-window effect (s)
2013	+95.2
2014	+42.4
2015	+184.2
2016	+132.3
2017	+105.4
2018	+78.8
2019	+238.2
2021	+151.9
2022	+319.7
2023	+170.3

Panel B: marathon summary and bounded sensitivities

Check	Effect (s)	95% CI	Detail
Equal-weight annual summary (10 events)	+151.8	[93.4, 210.3]	t(9), P = 0.0002
Incident-weighted active-window model	+151.1	[74.0, 228.2]	n = 108,222
Exclude source-flagged standby incidents	+151.5	[74.6, 228.5]	n = 108,064
Adjust initial call type and severity	+161.1	[63.9, 258.2]	n = 108,222
Incident-weighted prior-Saturday window (placebo)	+13.4	[-57.3, 84.2]	t(9), n = 108,222
Incident-weighted active minus prior Saturday	+137.7	[53.9, 221.4]	t(9), n = 108,222
Leave-one-event range			+133.2 to +164.0 s
Matched-date permutation	+151.8		P = 0.00005; 19,999 draws

Panel C: UNGA annual and summary estimates

Comparison	Effect (s)	95% CI
In-person year 2012	+117.7	[50.1, 185.4]
In-person year 2013	+92.6	[-57.0, 242.1]
In-person year 2014	+45.1	[-45.1, 135.3]
In-person year 2015	+227.9	[122.0, 333.8]
In-person year 2016	+123.2	[-46.5, 292.9]

Comparison	Effect (s)	95% CI
In-person year 2017	+93.9	[26.6, 161.2]
In-person year 2018	+133.3	[38.3, 228.3]
In-person year 2019	+55.1	[-34.2, 144.4]
In-person year 2021	+20.0	[-43.0, 82.9]
In-person year 2022	+192.9	[62.0, 323.8]
In-person year 2023	+137.3	[34.8, 239.8]
Pooled valid-response-weighted model	+113.1	[52.0, 174.3]
Equal-weight annual summary	+112.6	[71.2, 154.1]

Panel D: validity and scheduled-event stability

Check	Estimate	95% CI or range
Marathon active-window valid call-to-scene fraction (equal-weight annual)	-4.86	[-8.03, -1.69] pp
UNGA valid call-to-scene fraction	-1.55	[-3.56, 0.47] pp
Permit valid call-to-scene fraction	-0.54	[-1.39, 0.31] pp
Marathon active-window adverse stress test (278 invalid route-adjacent intervals at 1 s; 10/10 years positive)	+103.93	[60.9, 147.0] s
Marathon active-window raised validity ceilings (4,500-10,800 s; no assignment)		+164.1 to +251.4 s
Permit pooled call-to-scene	+37.44	[15.8, 59.1] s
Permit 2024	+42.80	[14.7, 70.9] s
Permit 2025	+30.82	[-4.0, 65.7] s
Permit leave-one-date range (187 omissions)		35.4 to 40.4 s

Supplementary Table S24. Frozen external-city replication of calendar endpoint-count directions. Panel A reports the same six positive-direction count contrasts and two cardiac-arrest comparison cells in New York City, Phoenix, Glendale and Cincinnati. Each external design fixed endpoint mappings, local-time windows and controls before estimation. Values are city-specific conditional Poisson IRRs and 95% confidence intervals under the frozen inference rule; no cross-city magnitude is pooled. The New York City column uses the unadjusted 2012-2023 estimates from the locked directional comparison and therefore differs slightly from the weather-adjusted headline estimates in Extended Data Tables 3 and 5. Phoenix and Glendale use medical Fire/EMS calls and a shared regional code vocabulary. Cincinnati includes all Fire/EMS incidents, so its all-call share diagnostic is not directly comparable and is not reported. Panel B reports the complete directional ledger. Sixteen of 18 external target cells and 22 of 24 target cells overall were positive; all eight cardiac-arrest comparison intervals included 1. A 17 October 2026 data audit found a Glendale source outage (23-29 October 2023, recorded as zero calls); a documented amendment to the frozen code excluded those days, which changed only the Glendale early-month ratio (from 1.097 to 1.090) and no direction (code 127).

Panel A: city-specific incidence-rate ratios [95% CI]

Schedule	Endpoint	New York City	Phoenix	Glendale	Cincinnati	Direction
Early month, days 1-3	Drug/poisoning	1.086 [1.072, 1.099]	1.020 [0.966, 1.076]	1.090 [0.947, 1.254]	1.035 [0.972, 1.101]	4/4 positive
July 4, 18:00-05:59	Assault	2.373 [1.952, 2.884]	1.448 [0.973, 2.154]	2.353 [1.359, 4.073]	1.560 [1.017, 2.394]	4/4 positive
July 4, 18:00-05:59	Burn	3.391 [2.512, 4.577]	5.000 [1.446, 17.286]	2.000 [0.027, 150.659]	6.400 [2.001, 20.467]	4/4 positive
July 4, 18:00-05:59	Cardiac arrest	1.071 [0.991, 1.158]	0.800 [0.271, 2.362]	1.500 [0.474, 4.752]	2.074 [0.829, 5.189]	4/4 intervals include 1
July 4, 18:00-05:59	Drug/poisoning	1.350 [1.233, 1.478]	1.401 [0.904, 2.173]	0.889 [0.125, 6.338]	1.725 [1.364, 2.182]	3/4 positive
Memorial/Labor Day	Assault	1.440 [1.298, 1.597]	1.381 [1.179, 1.619]	1.556 [0.956, 2.530]	1.261 [1.013, 1.570]	4/4 positive
Memorial/Labor Day	Cardiac arrest	1.006 [0.943, 1.073]	0.792 [0.492, 1.273]	0.828 [0.512, 1.337]	0.731 [0.470, 1.136]	4/4 intervals include 1
Memorial/Labor Day	Drug/poisoning	1.181 [1.111, 1.256]	0.960 [0.816, 1.129]	1.045 [0.620, 1.763]	1.050 [0.864, 1.277]	3/4 positive

Panel B: complete directional accounting

Metric	Result
External target cells with IRR > 1	16 / 18
All four-city target cells with IRR > 1	22 / 24
Target contrasts positive in all four cities	4 / 6
Target contrasts positive in three of four cities	2 / 6
Cardiac-arrest comparison intervals including 1	8 / 8

Supplementary Table S25. Frozen ZIP-hour resolution-parity analysis of calendar response-time contrasts. Before ZIP-level estimates were inspected, the three existing 2012-2023 response-time contrasts, localized fixed effects and +/-30-second decision threshold were frozen. Models are weighted by total calls and use ZIP-and-date-clustered inference. Panel A compares the localized estimates with the signed citywide results. The early-month interval lies within the frozen equivalence band, while both celebration estimates are below -30 seconds with intervals excluding zero; the result therefore met the full resolution-parity rule. Panel B reports exposure/control support and observed-mean coverage. The source lacks outcome-specific valid-response denominators, so this analysis addresses aggregation resolution but does not provide a causal or all-call response-time estimand.

Panel A: frozen resolution-parity estimates

Cell	ZIP-hour estimate (95% CI), s	Citywide estimate, s	Difference, s	Analytic ZIP-hours
Early month, days 1-3	+5.2 [-1.9, 12.3]	+5.3	-0.1	4,961,194
July 4	-67.1 [-93.9, -40.4]	-67.5	+0.4	123,929
Memorial/Labor Day	-92.9 [-123.5, -62.3]	-89.4	-3.5	252,131

Panel B: outcome-cell support

Cell	Exposure	Call-bearing ZIP-hours	Observed-mean ZIP-hours (coverage)	Total calls
Early month	Exposed	833,536	829,871 (99.56%)	1,510,662
Early month	Control	4,149,368	4,131,323 (99.57%)	7,470,572
July 4	Control	102,647	102,136 (99.50%)	188,112
July 4	Exposed	24,887	24,720 (99.33%)	44,602
Memorial/Labor Day	Control	203,951	203,060 (99.56%)	374,228
Memorial/Labor Day	Exposed	49,335	49,071 (99.46%)	88,171

Supplementary Table S26. Exploratory transport corroboration: design, support and complete diagnostics. The extension was fixed on 5-6 September 2026 after the parent permit-linked EMS association was known and before new trip outcomes were inspected. All 24 official monthly high-volume for-hire vehicle files from 2024-2025 were processed. Of 404 parent event occurrences, 17 were excluded before outcomes, 124 failed the fixed trip-support rules and 263 contributed to the primary result (166 dates; 74 exposed precincts). Failed events are retained in the execution ledger and do not enter as zero effects.

Taxi zones were assigned to the police precinct with the largest area overlap, requiring at least 50% overlap. Both trip endpoints had to lie in affected boroughs; either endpoint in an index-event precinct defined exposure. Integer pickup hours overlapped the half-open permit window. Controls used the same hours +/-7 and 14 days, excluding dates outside 2024-2025, daylight-saving transition dates and precincts carrying any full closure during the window. Actual comparison endpoints also had to be free of full closures. The mapping excludes precincts 105, 113 and 116, matching the parent specification. This measures endpoint-neighbourhood exposure without observing vehicle paths.

Eligible trips had positive duration below 7,200 s and positive distance no greater than 100 miles. Each retained origin-destination/hour cell required at least 5 eligible actual trips, 20 pooled control trips and positive control support on at least two dates. Events required 100 actual trips and 20 cells per arm, and at least two exposed endpoint zones. Cell event-minus-pooled-control means were weighted by control trip count divided by all eligible control dates, including dates with no eligible trips. Exposed and comparison changes were then differenced. No absent trip duration was imputed. The overall gate required at least 20 event dates, 15 exposed precincts and 20 distinct exposed origin-destination pairs. Fixed baseline weights address composition within retained cells, not missing, cancelled or low-support trips.

The primary summary gives each event equal weight. Date-cluster bootstrap intervals use 1,999 draws (seed 990605), retaining all events on each sampled date. The frozen corroboration rule required adequate coverage, a positive primary duration interval, positive estimates in both years and a pooled shifted-date placebo interval including zero with a smaller absolute point estimate than the primary. Pseudo-event pairs compare -14 with -7 days and +7 with +14 days. The placebo requires at least 20 trips on its single control date, whereas the primary pools controls; support therefore differs. The primary estimate restricted to the 50 placebo-supported events is a post-result comparison audit, not a replacement estimate.

EMS event contrasts use the same precincts, boroughs, hours and uncontaminated dates, with at least one actual valid response and valid responses on at least two control dates per precinct-hour. Fixed baseline weights use valid-response counts divided by all eligible control dates. The annual-source rebuild matched all 555,156 overlapping parent precinct-hours exactly. EMS and trip contrasts are paired only after their separate support rules; correlation is descriptive and does not identify mediation. TLC and EMS clocks were interpreted as New York civil time; all four 2024-2025 daylight-saving transition dates were excluded before estimation.

Panel A: all primary, year/type and placebo summaries

Sample	Measure	Unit	Estimate [95% interval]	Events (dates)
all	duration	seconds	+53.949 [43.283, 64.701]	263 (166)
all	miles	miles	+0.017 [0.009, 0.026]	263 (166)
all	count	pp	+0.614 [-0.679, 1.959]	263 (166)
2024	duration	seconds	+49.571 [34.556, 66.427]	109 (70)
2024	miles	miles	+0.019 [0.006, 0.033]	109 (70)
2024	count	pp	+0.905 [-0.881, 2.585]	109 (70)
2025	duration	seconds	+57.047 [41.643, 72.354]	154 (96)
2025	miles	miles	+0.016 [0.006, 0.026]	154 (96)
2025	count	pp	+0.409 [-1.414, 2.270]	154 (96)
Athletic races	duration	seconds	+47.149 [32.094, 64.879]	45 (37)
Athletic races	miles	miles	+0.031 [0.011, 0.054]	45 (37)
Athletic races	count	pp	+3.550 [0.898, 6.067]	45 (37)
Parade	duration	seconds	+55.352 [43.294, 68.893]	218 (140)
Parade	miles	miles	+0.014 [0.006, 0.023]	218 (140)
Parade	count	pp	+0.008 [-1.539, 1.479]	218 (140)
Pooled placebo	duration	seconds	-10.964 [-25.949, 2.548]	50 (42)
Pooled placebo	miles	miles	+0.003 [-0.011, 0.019]	50 (42)
Pooled placebo	count	pp	-7.801 [-11.372, -4.480]	50 (42)

The count contrast is the difference between exposed and comparison proportional changes in completed trips on retained cells. Its placebo is negative; it does not establish the absence of demand selection. The small mileage contrast does not support a large-detour interpretation. The leave-one-date duration mean ranges from +52.09 to +54.95 s.

Panel B: selected-trip population across supported events

Arm	Eligible trip appearances	Retained appearances	Aggregate retention (%)
Comparison	3,992,451	1,505,574	37.71
Exposed	2,104,873	483,743	22.98

Trip appearances can recur across overlapping event comparisons and are not unique citywide trips. Exact monthly URLs, queries, raw-file hashes, source schemas, the complete event ledger and failed support checks are preserved in the open-data module. The ordinary-transport result supports independent corroboration within this observed population.

Supplementary Table S27. Fire-company responses on marathon Sundays. The design was fixed before fitting. FDNY Fire Incident Dispatch Data (NYC Open Data 8m42-w767) were retrieved on 19 October 2026 for the 60 marathon and matched-control Sundays of 2012-2023 and their prior Saturdays, and joined to the corrected course classification by incident ZIP code (buffer ZIPs excluded). The outcome is the source incident response time with a valid flag and $0 < t < 3{,}600$ s. Each held race's event effect is the route-adjacent minus distant gap in cell means on the race Sunday minus the mean gap on its four control Sundays, the arithmetic of Supplementary Table S23; summaries average the ten races equally with t intervals (9 degrees of freedom). Non-medical incidents comprise non-medical emergencies, structural and non-structural fires and malfunction alarms; medical co-responses are fire-company responses to medical emergencies. Panel A compares the active-window event effect and own changes with the ambulance estimate; Panel B lists each race and the two placebo Sundays (2012, 2020; Supplementary Table S29). Fire companies may be pre-positioned for the event, and these records carry no event-standby flag.

Panel A: active window (07:00-20:59); ambulance event effect for comparison +151.8 s

Group	Quantity	Seconds (95% CI)	Races positive
Non-medical incidents	Route-adjacent minus distant (event effect)	+24.8 (4.0 to 45.7)	8 of 10
Non-medical incidents	Route-adjacent own change	+35.5 (13.6 to 57.4)	10 of 10
Non-medical incidents	Distant own change	+10.7 (-3.1 to 24.4)	5 of 10
Medical co-responses	Route-adjacent minus distant (event effect)	+29.7 (5.6 to 53.8)	8 of 10
Medical co-responses	Route-adjacent own change	+33.1 (3.5 to 62.7)	7 of 10
Medical co-responses	Distant own change	+3.4 (-10.3 to 17.1)	6 of 10
All fire incidents	Route-adjacent minus distant (event effect)	+25.1 (6.4 to 43.9)	8 of 10
All fire incidents	Route-adjacent own change	+33.6 (14.5 to 52.6)	9 of 10
All fire incidents	Distant own change	+8.4 (-4.7 to 21.6)	6 of 10

Panel B: active-window event effect by race

Race	Non-medical (s)	Medical co-response (s)
2012 (placebo)	-42.7	-77.7
2013	+15.1	-12.6
2014	-7.3	+8.7
2015	+56.1	-8.0
2016	+75.8	+76.8
2017	-15.2	+17.3

Race	Non-medical (s)	Medical co-response (s)
2018	+55.3	+45.3
2019	+10.0	+74.9
2020 (placebo)	-7.9	+57.5
2021	+27.1	+66.7
2022	+13.9	+8.3
2023	+17.5	+19.8

Supplementary Table S28. Comparison-trained transport sensitivity with matched placebo support. This single exploratory sensitivity was fixed on 7 September 2026 after the original transport result was known and before its own estimates were inspected. It retains the original events, endpoint geography, valid-trip definitions, civil clocks and closure exclusions (S26). The original five-event-trip cell minimum is removed. Four folds each hold out one of the +/-7 and 14-day comparison dates as a pseudo-event, training on the other three. Candidate origin-destination/hour cells require 20 valid training trips and positive trips on at least two eligible training dates. Baseline weights divide valid training trips by all eligible training dates, including zero-trip dates. Each arm must have 20 training cells and 100 baseline trips per day; exposed training cells must span at least two exposed zones. These support gates use comparison dates alone.

Event and held-out dates must both be geographically eligible and have at least one valid trip for a cell to contribute paired observed durations. Each contrast subtracts the training mean and uses the identical training weights on the identical cell intersection. Exposed and comparison changes are differenced. The common observed-cell requirement remains selection; no absent duration is zero-filled. There are no further event-day minima for trips, cells or zones. Supported folds are averaged equally within each event, then events equally. The 1,999 date-bootstrap draws (seed 1070907) are shared by the event, placebo and paired-difference estimates within each sample. The paired difference is a secondary descriptive contrast, not an identified closure effect.

Panel A: paired contrasts on the same supported population

Sample	Contrast	Seconds [95% interval]	Events (dates)
all	Event	+54.8 [40.8, 69.7]	248 (157)
all	Matched placebo	-4.3 [-10.8, 1.2]	248 (157)
all	Event minus placebo	+59.1 [44.4, 74.4]	248 (157)
2024	Event	+55.2 [35.0, 81.1]	101 (66)
2024	Matched placebo	-1.6 [-6.0, 2.4]	101 (66)
2024	Event minus placebo	+56.8 [35.5, 83.6]	101 (66)
2025	Event	+54.5 [35.4, 73.6]	147 (91)
2025	Matched placebo	-6.2 [-16.7, 2.3]	147 (91)
2025	Event minus placebo	+60.7 [40.8, 80.3]	147 (91)

All 387 planned events were processed; 842 of 1,548 folds and 248 events on 157 dates were supported. The retained exposed frame covers 74 precincts and 2,340 distinct origin-destination pairs. The other 139 events remain support failures. The frozen rule required complete sources, at least 20 dates, 15 exposed precincts and 20 exposed origin-destination pairs, a positive overall event interval, positive year estimates, and a placebo interval containing zero with a smaller absolute point estimate. The sensitivity meets that rule. This strengthens the completed-trip association and aligns the placebo population; it does not recover unobserved travel or isolate physical closure.

Panel B: support across paired folds; appearances can recur

Arm	Eligible appearances	Retained appearances	Trip retention (%)
Control	13,065,845	3,741,658	28.6
Exposed	7,133,302	1,135,036	15.9

Retention is calculated over supported event/fold/arm combinations. The paired cells retain 74.4% of exposed and 71.7% of comparison candidate training weight, but only 15.9% and 28.6% of eligible actual trip appearances. The candidate universe itself excludes many sparse cells; a high retained fraction of its weight is not high coverage of all trips. The full support CSV records candidate and shared cells, geographic ineligibility, zero-trip cells on each outcome date, lost baseline weight and definition failures for every fold, including unsupported folds.

Re-extraction reproduced every original support decision and duration contrast exactly, and all monthly cell and trip counts agreed. All 21 previously recorded full-file SHA-256 identities matched. For January-March 2024, the earlier range reads recorded no full-file SHA-256; their ETag, Last-Modified and Content-Length matched. This provenance feasibility correction was recorded before sensitivity estimation, with the failed check preserved. All 24 new full-file hashes and reduced monthly/event aggregates are now retained. Independent array-based reconstruction reproduced every fold, support count and bootstrap interval within 1e-8 seconds; the original analysis outputs are unchanged.

Supplementary Table S29. Clock-change placebo Sundays for the marathon design. Every held race fell on the Sunday daylight saving time ended. The design was fixed before running: the same extraction, phases, validity rule and event-effect arithmetic as Supplementary Table S23 were applied to 1 November 2020, a fall-back Sunday on which the race was virtual and no streets closed, and to 4 November 2012, a fall-back Sunday on which the race was cancelled after Hurricane Sandy, each with its four matched control Sundays and the corrected course classification. The same code reproduced the 2019 race effect to 1e-6. Panel A reports the placebo quantities by phase; Panel B reports the range across the ten held races. One pandemic-era Sunday and one storm-affected Sunday are descriptive checks, not tests.

Panel A: placebo Sundays, seconds

Placebo Sunday	Quantity	Prior Saturday	Setup	Active	Post
4 Nov 2012 (Sandy)	Route-adjacent minus distant (event effect)	-33.0	-94.3	-41.4	+11.2
4 Nov 2012 (Sandy)	Route-adjacent own change	+23.8	-35.5	+19.3	+18.4
4 Nov 2012 (Sandy)	Distant own change	+56.8	+58.8	+60.7	+7.2
1 Nov 2020 (virtual race)	Route-adjacent minus distant (event effect)	+16.6	-36.4	-1.7	-95.7
1 Nov 2020 (virtual race)	Route-adjacent own change	+128.8	+51.8	+22.9	-106.6
1 Nov 2020 (virtual race)	Distant own change	+112.2	+88.2	+24.6	-10.8

Panel B: ten held races, active window

Quantity	Mean (s)	Range (s)	Races positive
Distant own change	+43.0	-2.1 to +185.8	9 of 10
Route-adjacent own change	+194.9	+78.6 to +400.9	10 of 10
Route-adjacent minus distant (event effect)	+151.8	+42.4 to +319.7	10 of 10

Supplementary Table S30. Time of day of the early-month drug or poisoning increase. This exploratory extension was fixed on 28 September 2026, after the parent early-month result and an earlier descriptive day-1/day-16 hourly profile were known, and before any block model was fitted. It applies the primary design (days 1-3 versus same-weekday days 10-24, January excluded) separately to four local six-hour blocks, with month-year x weekday x hour matched-set fixed effects and a log physical-hour offset for daylight-saving transitions. The 2012-2023 model includes the published daily temperature. The 2024-2025 citywide counts lack weather and follow their own source eligibility, so magnitudes are not pooled across periods. Intervals use month-year clustering with t references. Panel A reports relative increases: every block is positive in both periods, with the largest point estimate in the afternoon in 2012-2023 and in the morning in 2024-2025. Panel B reports omnibus tests of equal block coefficients. Block differences are detected in 2012-2023 under both month-year and calendar-year clustering; the smaller 2024-2025 sample does not detect them, which does not establish equality. Panel C reports unadjusted matched differences in calls per physical hour. These are descriptive differences, not attributable cases, and no pairwise test of a single peak block was specified.

Panel A: relative increase on days 1-3 by six-hour block

Local hours	2012-2023, temperature adjusted (95% CI)	2024-2025, no weather (95% CI)
00:00-05:59	+4.2% (1.9 to 6.7)	+7.8% (0.6 to 15.4)
06:00-11:59	+7.6% (5.0 to 10.3)	+13.7% (5.5 to 22.5)
12:00-17:59	+13.1% (10.9 to 15.4)	+10.5% (4.9 to 16.5)
18:00-23:59	+8.2% (6.4 to 10.0)	+6.0% (1.5 to 10.6)

Panel B: omnibus tests of equal block coefficients

Model	Clusters	F (df)	P
2012-2023, temperature adjusted; month-year clusters	132	12.56 (3, 131)	2.88e-07
2012-2023, temperature adjusted; calendar-year clusters	12	4.62 (3, 11)	0.0251
2012-2023, no weather; month-year clusters	132	12.29 (3, 131)	3.88e-07
2024-2025, no weather; month-year clusters	22	1.78 (3, 21)	0.181
2024 only, no weather; month-year clusters	11	2.29 (3, 10)	0.14
2025 only, no weather; month-year clusters	11	0.46 (3, 10)	0.717

Panel C: unadjusted matched difference, calls per physical hour

Local hours	2012-2023 (95% CI)	2024-2025 (95% CI)
00:00-05:59	0.42 (0.16 to 0.68)	0.77 (0.01 to 1.53)
06:00-11:59	0.43 (0.26 to 0.59)	0.89 (0.37 to 1.41)
12:00-17:59	1.44 (1.19 to 1.68)	1.26 (0.59 to 1.94)
18:00-23:59	1.31 (1.00 to 1.61)	0.94 (0.24 to 1.64)

Supplementary Table S31. Joint calendar model of all pulses. The design was fixed and hashed on 15 October 2026 before fitting (design SHA-256 prefix b99ae74e), after the paper's separate pulse estimates and a 14 October exploratory decomposition were known. City-day counts for 2012-2023, January included, enter one conditional quasi-Poisson model per endpoint with year-month strata; day-of-week effects specific to calendar month; day-of-month terms for days 1-9 and 25-31 against days 10-24; indicators from 3 days before to 7 days after New Year's Day, July 4 (day 0 split into weekday and weekend years), Memorial Day and Labor Day; day-0 indicators for Christmas, Martin Luther King Jr. Day, Presidents' Day, Columbus Day and Thanksgiving; a temperature cross-basis (natural spline with 4 degrees of freedom, integer lags 0-3); precipitation; and an offset for physical day length. Intervals use year-month clustering (144 clusters) with t references. Panel C compares each holiday's day-0 coefficient with the same month's Sunday-minus-holiday-weekday coefficient; for weekday July 4 years, the comparator averages over the observed weekdays. Calendar-year-clustered sensitivities (12 clusters) produced non-positive-definite covariance corrections and could not be estimated reliably. Panel D tests departure from additive log-scale stacking when Labor Day falls on days 1-3, and compares the mean of days 1-3 with January excluded against the published 8.6%. Panel E fits the same model to ZIP-income-quartile counts with quartile-specific day-of-month and holiday day-0 terms. A base-R generalized linear model reproduced the fixest coefficients to less than 1e-14. Panel F reports overlap tests from a separate design fixed on 14 October before fitting, using the published matched designs with one added indicator: January days 1-3 and days 2-3 against other months; the day-4 excess in months other than July and the July 4 holiday-day ratio divided by it; July days 1-3 against other months; and September and November days 1-3 by whether Labor Day (2012-2014, 2018, 2019) or the marathon (2013-2015, 2019) fell within them. Detectable effects are 2.8 standard errors; gates reproduced the published holiday and early-month estimates exactly. The September and November rows include the holiday's or marathon's own excess (stacking); departure from additivity is in Panel D. The July 4 tail adjustment pools all 12 July 4s and uses the day-4 excess of other months; because July's early-month excess is itself larger (row 5), the adjusted value may still contain some of the tail. Weekday July 4s alone are in Panel B.

Panel A: day-of-month terms against days 10-24

Day of month	Drug or poisoning	Assault	Mental health	All calls
Day 25	-1.6% (-3.2 to -0.0)	-0.1% (-4.1 to 3.9)	-0.4% (-1.5 to 0.7)	-0.8% (-1.5 to -0.1)
Day 26	-2.1% (-3.7 to -0.5)	-4.0% (-8.0 to 0.3)	-1.6% (-3.0 to -0.2)	-0.9% (-1.6 to -0.2)
Day 27	-2.1% (-3.7 to -0.4)	+0.4% (-3.9 to 4.8)	-1.2% (-2.5 to 0.1)	-0.9% (-1.7 to -0.2)
Day 28	-2.0% (-3.6 to -0.3)	-0.1% (-4.4 to 4.4)	-1.1% (-2.2 to 0.0)	-1.1% (-1.7 to -0.4)
Day 29	-2.4% (-4.3 to -0.5)	-2.1% (-6.9 to 2.8)	-1.6% (-2.9 to -0.3)	-1.0% (-1.9 to -0.1)
Day 30	-0.4% (-2.5 to 1.7)	-0.4% (-4.5 to 3.8)	-1.6% (-3.1 to -0.2)	-0.8% (-1.6 to 0.1)
Day 31	+2.0% (-0.8 to 4.8)	+3.2% (-3.3 to 10.1)	-1.6% (-3.2 to 0.1)	-0.3% (-1.4 to 0.8)
Day 1	+8.2% (6.1 to 10.4)	+1.0% (-3.7 to 5.9)	+0.4% (-0.9 to 1.7)	+0.8% (-0.4 to 2.0)
Day 2	+8.9% (7.1 to 10.7)	-1.1% (-5.4 to 3.4)	-0.2% (-1.4 to 1.1)	+1.2% (0.1 to 2.4)

Day of month	Drug or poisoning	Assault	Mental health	All calls
Day 3	+7.6% (5.7 to 9.4)	-1.8% (-5.9 to 2.5)	-0.3% (-1.6 to 0.9)	+1.1% (-0.0 to 2.2)
Day 4	+6.9% (5.1 to 8.7)	-3.9% (-8.2 to 0.6)	+1.3% (0.1 to 2.6)	+1.7% (0.6 to 2.8)
Day 5	+4.9% (3.2 to 6.7)	-4.0% (-8.2 to 0.4)	-0.0% (-1.3 to 1.2)	+0.8% (-0.2 to 1.9)
Day 6	+3.0% (1.0 to 4.9)	-5.1% (-9.3 to -0.7)	+1.3% (0.1 to 2.4)	+1.0% (0.1 to 2.0)
Day 7	+3.0% (1.4 to 4.7)	-7.1% (-11.1 to -2.9)	+0.6% (-0.6 to 1.8)	+0.4% (-0.5 to 1.3)
Day 8	+2.1% (0.5 to 3.7)	-3.6% (-7.8 to 0.8)	+1.4% (0.3 to 2.5)	+0.2% (-0.6 to 1.0)
Day 9	+1.7% (0.2 to 3.1)	-4.0% (-7.9 to 0.0)	-0.0% (-1.2 to 1.2)	+0.2% (-0.5 to 1.0)

Panel B: holiday lead and lag terms (days relative to the holiday)

Term	Drug or poisoning	Assault	Mental health	All calls
New Year's Day -3	+1.4% (-3.6 to 6.7)	+11.8% (-0.8 to 25.9)	-4.0% (-7.0 to -0.9)	+0.1% (-2.3 to 2.6)
New Year's Day -2	+5.9% (1.9 to 10.1)	-6.1% (-22.9 to 14.5)	-2.0% (-5.4 to 1.7)	-0.9% (-3.2 to 1.5)
New Year's Day -1	+28.9% (20.3 to 38.3)	+16.1% (-1.8 to 37.2)	-7.1% (-11.4 to -2.7)	-2.5% (-4.6 to -0.3)
New Year's Day 0	+102.0% (75.9 to 132.0)	+137.8% (92.3 to 194.1)	-15.2% (-19.2 to -11.1)	+16.8% (11.8 to 22.1)
New Year's Day +1	-8.6% (-16.9 to 0.4)	-2.3% (-17.7 to 16.0)	-6.9% (-11.4 to -2.2)	-0.1% (-4.1 to 4.0)
New Year's Day +2	-7.5% (-12.6 to -2.2)	+8.1% (-9.7 to 29.5)	-3.1% (-6.3 to 0.1)	+1.9% (-1.9 to 5.9)
New Year's Day +3	-6.7% (-12.0 to -1.0)	+15.3% (1.2 to 31.4)	-4.5% (-9.2 to 0.4)	-0.0% (-3.4 to 3.5)
New Year's Day +4	-3.0% (-8.1 to 2.3)	+2.2% (-12.8 to 19.9)	-2.5% (-6.1 to 1.2)	+4.0% (-0.7 to 8.8)
New Year's Day +5	-0.0% (-4.1 to 4.3)	+5.1% (-4.0 to 15.1)	-3.4% (-6.7 to 0.0)	+2.6% (-0.1 to 5.3)
New Year's Day +6	+1.6% (-2.1 to 5.4)	+12.6% (0.7 to 25.9)	+0.1% (-3.5 to 3.8)	+1.9% (-0.7 to 4.6)
New Year's Day +7	-0.6% (-4.5 to 3.5)	-8.6% (-23.3 to 9.0)	-3.2% (-7.2 to 1.0)	+0.8% (-2.0 to 3.6)
July 4 -3	+3.7% (-1.4 to 9.0)	+3.0% (-11.5 to 19.7)	-1.0% (-4.7 to 2.8)	+1.6% (-1.1 to 4.3)
July 4 -2	+2.2% (-3.9 to 8.6)	+10.3% (-6.0 to 29.6)	+3.9% (-0.3 to 8.2)	+1.7% (-0.2 to 3.7)
July 4 -1	+10.2% (5.4 to 15.3)	+11.1% (-3.6 to 28.1)	-1.9% (-4.2 to 0.5)	+0.1% (-1.9 to 2.2)
July 4 0 (weekday years)	+22.8% (16.9 to 28.9)	+83.4% (61.4 to 108.3)	-17.8% (-20.3 to -15.2)	-7.0% (-8.7 to -5.3)
July 4 0 (weekend years)	-2.4% (-8.4 to 4.0)	+20.2% (11.0 to 30.2)	-7.6% (-13.6 to -1.2)	-1.2% (-4.6 to 2.4)
July 4 +1	+11.0% (5.0 to 17.5)	+81.8% (60.4 to 106.2)	+0.8% (-3.8 to 5.6)	+6.1% (3.3 to 9.1)
July 4 +2	-3.8% (-7.6 to 0.2)	-1.5% (-17.9 to 18.3)	-2.1% (-6.2 to 2.1)	-1.8% (-3.7 to 0.2)
July 4 +3	-4.2% (-9.0 to 0.7)	+9.2% (-2.2 to 22.1)	-0.2% (-4.0 to 3.7)	+0.0% (-1.9 to 2.0)
July 4 +4	-3.1% (-8.3 to 2.5)	-5.8% (-16.9 to 6.7)	+2.2% (-1.9 to 6.4)	-1.4% (-2.9 to 0.0)
July 4 +5	+1.0% (-2.7 to 4.8)	+5.7% (-7.6 to 21.0)	+3.2% (-0.7 to 7.2)	-0.6% (-2.2 to 1.2)
July 4 +6	+6.2% (2.3 to 10.3)	-4.6% (-16.3 to 8.7)	+3.1% (-0.5 to 6.8)	+1.4% (-0.1 to 3.0)
July 4 +7	+2.5% (-1.7 to 6.9)	+4.6% (-6.1 to 16.4)	+1.3% (-0.7 to 3.3)	+0.5% (-0.7 to 1.6)
Memorial Day -3	-3.4% (-6.3 to -0.3)	+2.1% (-15.1 to 22.8)	+1.0% (-2.4 to 4.4)	-0.9% (-2.7 to 1.0)
Memorial Day -2	-0.5% (-4.8 to 4.0)	+0.6% (-13.7 to 17.3)	-2.6% (-7.3 to 2.3)	-2.7% (-5.2 to -0.2)
Memorial Day -1	+2.5% (-0.5 to 5.7)	+7.8% (-2.7 to 19.3)	-1.1% (-5.3 to 3.3)	-2.3% (-4.6 to 0.0)
Memorial Day 0	+7.6% (2.0 to 13.6)	+23.9% (7.6 to 42.6)	-19.3% (-22.1 to -16.4)	-10.2% (-12.4 to -8.0)
Memorial Day +1	-0.9% (-6.1 to 4.7)	+7.9% (-2.3 to 19.1)	-0.2% (-5.0 to 4.9)	+1.7% (-0.2 to 3.5)
Memorial Day +2	-6.2% (-12.3 to 0.4)	-0.4% (-10.9 to 11.3)	+0.5% (-3.0 to 4.1)	-3.2% (-5.1 to -1.2)
Memorial Day +3	-2.6% (-8.7 to 3.8)	-8.2% (-17.9 to 2.6)	-0.2% (-3.6 to 3.3)	-1.7% (-4.1 to 0.7)

Term	Drug or poisoning	Assault	Mental health	All calls
Memorial Day +4	-4.4% (-7.5 to -1.2)	-4.2% (-16.9 to 10.5)	+0.0% (-2.7 to 2.8)	-0.1% (-1.3 to 1.2)
Memorial Day +5	-1.5% (-5.3 to 2.5)	+0.2% (-8.3 to 9.6)	-0.8% (-4.2 to 2.7)	+1.2% (-1.2 to 3.6)
Memorial Day +6	-3.0% (-6.5 to 0.5)	-4.6% (-15.2 to 7.2)	+1.4% (-2.8 to 5.6)	-0.1% (-2.6 to 2.5)
Memorial Day +7	-0.7% (-5.5 to 4.3)	+8.1% (-1.0 to 17.9)	+2.7% (-2.4 to 8.0)	+1.2% (-0.9 to 3.3)
Labor Day -3	+4.5% (-1.2 to 10.5)	-1.8% (-14.4 to 12.7)	-4.2% (-8.3 to 0.1)	-3.0% (-5.0 to -1.0)
Labor Day -2	+3.1% (-3.0 to 9.5)	+1.6% (-6.7 to 10.6)	+1.7% (-3.1 to 6.7)	-0.6% (-2.8 to 1.7)
Labor Day -1	+2.5% (-2.8 to 8.0)	+8.3% (-3.1 to 21.0)	-2.7% (-6.5 to 1.2)	-3.5% (-5.9 to -1.0)
Labor Day 0	+22.3% (16.2 to 28.8)	+66.1% (50.2 to 83.7)	-14.0% (-18.2 to -9.7)	-7.8% (-9.8 to -5.6)
Labor Day +1	-2.8% (-8.4 to 3.1)	+11.1% (-4.1 to 28.8)	-4.6% (-9.2 to 0.3)	-2.0% (-3.8 to -0.1)
Labor Day +2	-5.2% (-9.4 to -0.9)	-5.3% (-16.1 to 6.8)	-5.6% (-7.8 to -3.3)	-5.3% (-7.4 to -3.2)
Labor Day +3	-1.3% (-7.2 to 5.0)	+5.3% (-6.3 to 18.4)	-3.2% (-6.1 to -0.2)	-3.5% (-5.2 to -1.7)
Labor Day +4	-1.3% (-7.3 to 5.1)	+5.7% (-6.3 to 19.2)	-2.0% (-6.7 to 2.9)	-4.2% (-5.7 to -2.7)
Labor Day +5	-0.7% (-6.5 to 5.5)	+1.6% (-10.1 to 14.7)	-0.4% (-5.0 to 4.4)	-1.6% (-3.5 to 0.3)
Labor Day +6	-1.3% (-6.2 to 3.7)	-11.7% (-22.1 to 0.2)	-0.6% (-6.0 to 5.1)	-3.8% (-5.7 to -1.8)
Labor Day +7	+4.1% (-1.3 to 9.7)	+2.8% (-12.0 to 20.1)	-0.2% (-4.8 to 4.6)	-1.5% (-3.5 to 0.5)

Panel C: holidays against the same month's Sunday pattern

Holiday	Endpoint	Holiday vs same weekday	Ordinary Sunday vs same weekday	Holiday beyond Sunday
Labor Day	All calls	-7.8% (-9.8 to -5.6)	-5.4% (-6.5 to -4.2)	-2.5% (-4.3 to -0.8)
Labor Day	Assault	+66.1% (50.2 to 83.7)	+40.7% (28.4 to 54.1)	+18.1% (3.4 to 34.8)
Labor Day	Drug or poisoning	+22.3% (16.2 to 28.8)	+28.0% (23.6 to 32.7)	-4.5% (-8.5 to -0.2)
Labor Day	Mental health	-14.0% (-18.2 to -9.7)	-15.3% (-17.7 to -12.9)	+1.5% (-3.6 to 6.9)
Memorial Day	All calls	-10.2% (-12.4 to -8.0)	-7.6% (-9.4 to -5.7)	-2.8% (-4.9 to -0.8)
Memorial Day	Assault	+23.9% (7.6 to 42.6)	+40.8% (27.3 to 55.7)	-12.0% (-24.5 to 2.5)
Memorial Day	Drug or poisoning	+7.6% (2.0 to 13.6)	+15.1% (10.6 to 19.8)	-6.5% (-9.3 to -3.5)
Memorial Day	Mental health	-19.3% (-22.1 to -16.4)	-18.2% (-21.6 to -14.6)	-1.4% (-5.3 to 2.7)
Memorial and Labor Day	All calls	-9.0% (-10.4 to -7.5)	-6.5% (-7.6 to -5.4)	-2.7% (-3.9 to -1.4)
Memorial and Labor Day	Assault	+43.4% (31.5 to 56.4)	+40.7% (31.5 to 50.6)	+1.9% (-7.8 to 12.7)
Memorial and Labor Day	Drug or poisoning	+14.7% (10.6 to 19.0)	+21.4% (18.2 to 24.7)	-5.5% (-7.9 to -3.0)
Memorial and Labor Day	Mental health	-16.7% (-19.2 to -14.2)	-16.8% (-18.9 to -14.6)	+0.1% (-3.2 to 3.4)
Weekday July 4	All calls	-7.0% (-8.7 to -5.3)	-3.7% (-4.6 to -2.7)	-3.5% (-5.6 to -1.3)
Weekday July 4	Assault	+83.4% (61.4 to 108.3)	+33.2% (24.6 to 42.5)	+37.6% (20.5 to 57.2)
Weekday July 4	Drug or poisoning	+22.8% (16.9 to 28.9)	+16.6% (13.7 to 19.5)	+5.3% (0.0 to 10.8)
Weekday July 4	Mental health	-17.8% (-20.3 to -15.2)	-14.9% (-17.1 to -12.7)	-3.3% (-6.8 to 0.3)

Panel D: coinciding pulses and bridge to the published early-month estimate

Quantity	Drug or poisoning	Assault	Mental health	All calls
Labor Day on days 1-3, departure from additivity	+3.7% (-3.1 to 11.0)	+9.9% (-3.9 to 25.6)	+2.3% (-7.0 to 12.5)	-0.3% (-3.8 to 3.3)
Mean of days 1-3, January excluded	+8.2% (6.8 to 9.6)	-0.5% (-3.5 to 2.6)	-0.1% (-1.0 to 0.9)	+1.1% (-0.0 to 2.1)

Panel E: drug or poisoning terms by ZIP income quartile

Term	Lowest income	Q2	Q3	Highest income
Day 2	+11.9% (9.6 to 14.3)	+9.2% (6.6 to 11.9)	+6.3% (2.9 to 9.9)	+4.8% (2.0 to 7.8)
Day 25	-0.2% (-2.1 to 1.8)	-0.3% (-2.7 to 2.2)	-1.5% (-4.4 to 1.5)	-6.2% (-9.9 to -2.3)
Day 28	-1.2% (-3.3 to 0.9)	-1.6% (-3.9 to 0.8)	-0.9% (-3.9 to 2.3)	-4.8% (-7.9 to -1.5)
Day 9	+2.9% (0.8 to 5.1)	+2.8% (0.7 to 5.0)	+0.2% (-2.6 to 3.2)	-0.8% (-3.4 to 1.9)
July 4 0 (weekday years)	+22.3% (12.3 to 33.2)	+24.4% (15.5 to 33.9)	+31.9% (21.2 to 43.6)	+14.7% (7.3 to 22.6)
Labor Day 0	+19.6% (10.4 to 29.6)	+39.7% (25.3 to 55.9)	+16.4% (6.6 to 27.0)	+10.5% (2.8 to 18.7)
Mean of days 1-3	+11.0% (9.3 to 12.7)	+7.6% (5.7 to 9.5)	+6.1% (3.8 to 8.4)	+5.7% (3.5 to 8.1)
Memorial Day 0	+7.8% (1.4 to 14.6)	+14.8% (3.8 to 27.0)	+11.1% (-2.1 to 26.0)	-6.4% (-19.1 to 8.2)
New Year's Day 0	+84.0% (61.6 to 109.6)	+89.2% (70.0 to 110.7)	+99.3% (68.4 to 136.0)	+159.0% (113.4 to 214.5)

Panel E (continued): joint tests across income quartiles

Terms tested equal across quartiles	F (df)	P
Days 1-9	3.14 (27, 143)	5.73e-06
Days 25-31	0.97 (21, 143)	0.504
Days 1-3	4.54 (9, 143)	2.9e-05

Panel F: overlapping pulses (14 October design, fixed before fitting)

Test	Drug or poisoning	All calls	Support	Minimum detectable drug or poisoning effect
January days 1-3 vs other months (New Year's carryover)	+30.1% (12.5 to 50.4)	+6.1% (2.4 to 9.8)	36 January case days	23.0%
January days 2-3 vs other months	-5.7% (-11.1 to 0.0)	+0.4% (-2.1 to 3.1)	24 January case days	8.8%
Day 4 vs mid-month, months other than July	+6.5% (4.7 to 8.3)	+1.7% (0.7 to 2.7)	120 day-4 strata	2.4%
July 4 full day beyond the day-4 tail (unadjusted +22.3%)	+14.9% (6.3 to 24.1)		July 4 unadjusted +22.3%; day-4 +6.5%	11.7%
July days 1-3 vs other months	+7.2% (3.7 to 10.9)	+0.4% (-0.9 to 1.8)	12 Julys	4.9%
September days 1-3 with vs without Labor Day	+9.2% (1.8 to 17.1)	-1.8% (-4.8 to 1.4)	5 vs 7 Septembers	10.5%

Test	Drug or poisoning	All calls	Support	Minimum detectable drug or poisoning effect
Labor Day on days 1-3 vs days 4-7	+9.0% (-0.3 to 19.1)	+0.7% (-4.5 to 6.3)	5 vs 7 Labor Days	13.6%
November days 1-3 with vs without the marathon	+5.0% (-3.2 to 13.9)	-0.1% (-3.4 to 3.3)	4 vs 8 Novembers	12.3%

Supplementary Table S32. Pulses across 170 ZIPs: two-stage multivariate meta-analysis. Stage 1 fits, in each ZIP, a Poisson model of daily drug or poisoning calls with year-month strata, day-of-week effects, days 1-3 (EM), days 4-9, days 25-31 (LATE), Memorial and Labor Day (ML), July 4 (J4), New Year's Day, other federal holidays, a temperature spline, precipitation and a day-length offset, with year-month clustering. Estimates with absolute value above 5 or standard error above 5 (separation) were set missing; one ZIP lacked an ML estimate. Stage 2 pools the four-parameter vectors with restricted maximum likelihood (mixmeta), giving pooled effects, between-ZIP standard deviations (tau) and the correlation of true ZIP effects across pulses; correlation intervals come from 499 ZIP bootstrap resamples. The ML heterogeneity test is univariate because mixmeta does not return an outcome-specific Q when one outcome is missing for a ZIP. Moran's I uses best linear unbiased predictions and queen contiguity of DOHMH modified ZCTA polygons (168 units) with 999 permutations. Panel D converts predictions into excess drug or poisoning calls per 10,000 residents per year from the four pulses; it is descriptive. metafor reproduced the between-ZIP correlations to less than 3e-5.

Panel A: pooled ZIP effects and heterogeneity

Parameter	Pooled (95% CI)	Between-ZIP SD (log scale)	Heterogeneity P
Days 1-3	+7.7% (6.6 to 8.8)	0.044	2.17e-10
Days 25-31	-0.9% (-1.5 to -0.3)	0.021	0.00111
Memorial and Labor Day	+14.2% (9.4 to 19.2)	0.205	< 1e-15
July 4	+22.6% (17.8 to 27.6)	0.147	2.11e-05

Panel B: between-ZIP correlations across pulses

Pulse pair	Correlation of true ZIP effects	Bootstrap 95% interval	Resamples
EM with J4	-0.20	-0.60 to 0.23	499
EM with LATE	0.34	-0.14 to 0.73	499
EM with ML	0.18	-0.19 to 0.47	499
LATE with J4	-0.11	-0.75 to 0.46	499
LATE with ML	0.10	-0.35 to 0.66	499
ML with J4	0.42	-0.06 to 0.89	499

Panel C: spatial clustering of ZIP effects

Parameter	Moran's I	Permutation P	Units
Days 1-3	0.38	0.001	168
Days 25-31	0.11	0.026	168
Memorial and Labor Day	0.29	0.001	168
July 4	0.10	0.037	168

Panel D: descriptive annual excess from the four pulses

Income quartile	ZIPs	Median per 10,000 per year	Interquartile range	Total excess calls per year
Q1	46	0.97	0.47 to 1.86	331
Q2	44	0.62	0.39 to 1.12	176
Q3	46	0.34	0.17 to 0.58	88
Q4	33	0.37	0.15 to 1.03	70

Supplementary Table S33. Neighbourhood moderators of the ZIP pulse effects. An amendment fixed and hashed before any moderator data were downloaded (15 October 2026; design SHA-256 prefix e4f3852c) extended stage 1 to five parameters by also extracting New Year's Day (NY), with the stage-1 model unchanged. Moderators were standardized across ZIPs: log median income and race and ethnicity from the American Community Survey 2015-2019 (table B03002: non-Hispanic Black and Hispanic shares); log population density from DOHMH modified ZCTA polygons; NYCHA residents allocated to ZIPs by each development's address share (216 of 346 developments, 284,658 residents, matched to addresses); active Department of Consumer and Worker Protection pawnbroker licences per 100,000 residents (licensed check cashers have no open list); active on-premises State Liquor Authority licences per 10,000 residents; overcrowded housing; and subway stations per 100,000 residents. Licence and facility lists are current snapshots used as structural proxies for 2012-2023. Panel A tests each moderator jointly across the five pulses with Benjamini-Hochberg control. Panel B fits income and both race and ethnicity shares together. Panel C reports four prespecified directional hypotheses, each adjusted for income. Panel D shows the New Year's income slope with and without alcohol-licence density. All measures are ecological and attach to incident location, not residence; racial composition marks residential segregation, not individual risk. metafor reproduced the Panel B slopes. Panel E is a further amendment fixed and hashed on 19 October before benefit data were pulled (design SHA-256 prefix 5bc2a7b3): the Panel B model with borough fixed effects, with ACS 2015-2019 household shares receiving Supplemental Security Income, cash public assistance and SNAP, and with both. Its prespecified rule called the early-month Black-share slope robust if every interval excluded zero and retained at least half of the Panel B estimate and a within-borough permutation test (999 permutations of Black share, univariate REML meta-regression) gave $P < 0.05$; the rule was met. Residual Moran's I uses the Panel C geometry. Benefit shares are strongly collinear with log income (r -0.78 to -0.89). One standard deviation across ZIPs is 23.8 percentage points of non-Hispanic Black share, 19.4 points of Hispanic share and 0.44 log units (a 1.56-fold difference) of median income. metafor reproduced the Panel E slopes to less than 1e-6.

Panel A: joint tests across the five pulses

Moderator	Chi-squared (5 df)	P	Benjamini-Hochberg q
log median income	30.4	1.26e-05	4.67e-05
% non-Hispanic Black	80.3	< 1e-15	6.41e-15
% Hispanic	5.9	0.319	0.359
log population density	29.9	1.56e-05	4.67e-05
% NYCHA residents	19.4	0.0016	0.00361
pawnbrokers per 100k	14.0	0.0158	0.0285
on-premises alcohol licences per 10k	10.4	0.0653	0.0979
% overcrowded	2.0	0.85	0.85
subway stations per 100k	9.3	0.0969	0.125

Panel B: income and race and ethnicity in one model, % per SD (race block Wald P = 2.46e-11)

Term	Days 1-3	Days 25-31	Memorial and Labor Day	July 4	New Year's Day
% Hispanic	+1.0% (-0.1 to 2.1)	+0.4% (-0.5 to 1.3)	-2.9% (-8.1 to 2.6)	-0.1% (-5.6 to 5.7)	+6.5% (0.5 to 12.8)
% non-Hispanic Black	+3.4% (2.3 to 4.5)	+0.5% (-0.3 to 1.4)	+7.8% (2.3 to 13.5)	+0.2% (-5.0 to 5.7)	-4.5% (-9.5 to 0.9)
Log median income	-0.0% (-1.2 to 1.2)	-0.0% (-1.0 to 1.0)	-2.9% (-8.7 to 3.3)	-0.2% (-6.2 to 6.2)	+11.7% (4.7 to 19.1)

Panel C: prespecified hypotheses, adjusted for income

Hypothesis	Per SD (95% CI)	P	ZIPs
pawnbrokers -> early month (+)	-0.3% (-1.0 to 0.4)	0.402	169
NYCHA share -> early month (+)	+1.2% (0.2 to 2.1)	0.0142	169
on-premises alcohol -> New Year's (+)	+56.6% (19.9 to 104.6)	0.00101	169
on-premises alcohol -> July 4 (+)	-30.7% (-45.2 to -12.4)	0.00215	169

Panel D: New Year's income gradient and alcohol-licence density

Model	New Year's income slope per SD (95% CI)	P
income only	+8.8% (4.4 to 13.5)	6.55e-05
income + on-premises alcohol	+4.6% (-0.1 to 9.6)	0.0575

Panel E: robustness of the early-month slopes, % per SD

Model	% non-Hispanic Black	log median income	% Hispanic	% households with SSI	% with cash public assistance	% receiving SNAP	Within-borough permutation P	Residual Moran's I (P)
Panel B model	+3.4% (2.3 to 4.5)	-0.0% (-1.2 to 1.2)	+1.0% (-0.1 to 2.1)					0.090 (0.054)
+ borough	+3.3% (2.2 to 4.4)	+0.1% (-1.1 to 1.3)	+0.8% (-0.4 to 2.1)				0.001	
+ benefit receipt	+3.3% (2.2 to 4.4)	+2.0% (0.0 to 4.0)	+0.9% (-0.3 to 2.1)	+2.2% (0.2 to 4.1)	+1.6% (-0.0 to 3.3)	-1.2% (-3.6 to 1.3)		
+ borough and benefit receipt	+3.2% (2.1 to 4.3)	+1.0% (-1.0 to 3.1)	+1.0% (-0.3 to 2.4)	+1.8% (-0.3 to 3.8)	+1.6% (-0.1 to 3.3)	-2.3% (-4.8 to 0.3)	0.001	0.045 (0.146)

Supplementary Table S34. Marathon delay across space. Panel A splits the headline equal-event estimate into route-adjacent (within 1 km) and distant (more than 5 km) changes, each the event-Sunday mean minus the mean of the four matched control Sundays among incidents with a valid call-to-scene interval, by phase. The difference reproduces the headline +151.84 s exactly; t intervals use 9 degrees of freedom. Panel B uses ZIP-hour call-to-scene means weighted by total calls for active hours on the same dates, by distance band; this estimator has no validity filter and is not comparable in magnitude with Panel A. Panel C records the prespecified street-network test: no distant ZIP's nearest emergency-department hospital (54 New York State Department of Health facilities with emergency-department certification and coordinates) lay across the route from the ZIP centroid, so the test had no variation and was not estimated. A patient's nearest hospital does not identify where the responding ambulance started, so this test could not have isolated a call-to-scene mechanism. The course was reconstructed along its documented streets (42.5 km; Supplementary Methods). A pandas recomputation matched Panel A to less than 1e-12.

Panel A: incident-level changes from control Sundays, seconds

Phase	Route-adjacent change	Distant change	Difference (headline estimator)
Saturday 21:00-23:59	+24.6 (-75.6 to 124.7); 6/10 positive	+20.8 (-34.3 to 75.9); 7/10 positive	+3.7 (-79.4 to 86.9); 6/10 positive
Sunday 00:00-06:59	+37.8 (-42.5 to 118.1); 6/10 positive	+34.2 (-42.1 to 110.5); 5/10 positive	+3.5 (-67.0 to 74.0); 3/10 positive
Sunday 07:00-20:59	+194.9 (120.3 to 269.4); 10/10 positive	+43.0 (2.8 to 83.3); 9/10 positive	+151.8 (93.4 to 210.3); 10/10 positive
Sunday 21:00-23:59	+30.6 (-57.8 to 119.0); 5/10 positive	-16.2 (-36.3 to 3.9); 2/10 positive	+46.8 (-34.2 to 127.8); 5/10 positive

Panel B: ZIP-hour active-window changes by distance

Distance from route	Change, s (95% CI)	Positive races	ZIPs
<=1 km	+286.7 (124.0 to 449.4)	10/10	24
1-3 km	+114.4 (47.3 to 181.5)	9/10	37
3-5 km	+103.1 (-5.3 to 211.5)	10/10	27
>5 km	+72.5 (-23.0 to 168.0)	10/10	82

Panel C: street-network test support

Distant ZIPs evaluated	Nearest hospital across route	Median distance to nearest hospital, km
82	0	2.5

Supplementary Table S35. The joint calendar structure in Phoenix, Glendale and Cincinnati. A second amendment, fixed and hashed before fitting (17 October 2026; design SHA-256 prefix 852acc32), applied the Supplementary Table S31 joint model to daily counts from the frozen external-city scripts, evaluated only up to panel construction so that their source checksums, endpoint crosswalks and validation gates applied unchanged; per-year totals matched the published source audits exactly. The external panels lack precipitation, so that covariate is omitted. A scoped data audit (17 October 2026) found a Glendale source outage (23-29 October 2023, recorded as zero calls); days with zero total calls are excluded as outages. Model days also require temperature at lags 0-3, which removes a further 48 Glendale days. Phoenix temperature begins in December 2015, so the primary Phoenix model covers 2016-2018; a no-temperature sensitivity uses all five years. Prespecified predictions: days 1-3 above, and days 25-29 below, mid-month levels for drug or poisoning calls; Memorial and Labor Day not above the same month's Sunday pattern; weekday July 4 assault above it. Intervals use year-month clustering with t references; several fits required covariance corrections, so intervals are approximate. Magnitudes are not pooled across cities. A base-R generalized linear model reproduced the Cincinnati coefficients to less than 1e-15.

Panel A: model samples

City	Model days	First and last date	Outage days excluded	Memorial / Labor / weekday July 4
Phoenix	1124	2015-12-04 to 2018-12-31	0	3 / 3 / 3
Phoenix (no temperature, 2014-2018)	1826	2014-01-01 to 2018-12-31	0	5 / 5 / 4
Glendale	1768	2021-01-04 to 2025-12-31	7	5 / 5 / 4
Cincinnati	4014	2015-01-04 to 2025-12-30	0	11 / 11 / 8

Panel B: prespecified predictions (* = not in the predicted direction)

Prediction	Phoenix	Phoenix (no temperature, 2014-2018)	Glendale	Cincinnati
Days 1-3 above mid-month (drug or poisoning)	+2.6% (-5.3 to 11.1)	-0.5% (-6.7 to 6.1) *	+8.4% (-7.0 to 26.5)	+4.4% (-2.7 to 12.0)
Days 25-29 below mid-month (drug or poisoning)	-0.8% (-6.3 to 5.2)	-1.6% (-5.9 to 2.8)	-12.8% (-21.9 to -2.6)	-1.0% (-6.7 to 5.0)
Memorial/Labor not above Sunday (assault)	+11.9% (-9.0 to 37.6)	+8.7% (-7.2 to 27.4)	-4.3% (-26.7 to 24.9)	-13.6% (-26.9 to 2.0)
Memorial/Labor not above Sunday (drug or poisoning)	+8.2% (-7.2 to 26.1)	+6.2% (-7.4 to 21.9)	+8.5% (-29.4 to 66.8)	+11.7% (-9.8 to 38.4)
Weekday July 4 assault above Sunday	+9.5% (-7.5 to 29.8)	+18.6% (2.2 to 37.6)	-29.1% (-58.8 to 22.1) *	-13.5% (-35.7 to 16.3) *

Panel C: all calls on Memorial and Labor Day against the Sunday pattern

Quantity	Cincinnati	Glendale	Phoenix	Phoenix (no temperature, 2014-2018)
Memorial/Labor beyond Sunday	-2.4% (-5.6 to 0.9)	-6.4% (-15.0 to 3.1)	-7.2% (-11.6 to -2.5)	-4.1% (-6.4 to -1.8)
Memorial/Labor vs same weekday	-7.6% (-10.6 to -4.5)	-12.9% (-19.7 to -5.6)	-14.4% (-17.5 to -11.2)	-11.4% (-13.9 to -8.9)
Sunday vs same weekday	-5.3% (-6.9 to -3.7)	-7.0% (-12.5 to -1.2)	-7.8% (-10.0 to -5.5)	-7.6% (-9.4 to -5.8)

Panel D: similarity of the drug or poisoning day-of-month curve to New York

City	Day-of-month terms	Correlation with New York curve
Phoenix	16	0.19
Phoenix (no temperature, 2014-2018)	16	0.11
Glendale	16	0.59
Cincinnati	16	0.57

Supplementary Table S36. Weekly-coupling dose-response for Memorial and Labor Day in New York City. Fixed in the same amendment. For each of the 24 holidays and each endpoint, the holiday deviation is the log ratio of the holiday count to the mean of the Mondays 7 days before and after; the period's Sunday contrast is the log ratio of the mean of the Sundays 15 days before and 13 days after to the mean of the Mondays 14 days before and after. The two use no common control days, so shared noise cannot create the association. A slope of 1 would mean the holiday reproduces its period's Sunday-versus-Monday contrast in full. Slopes are ordinary least squares with HC2 standard errors and t intervals, with Benjamini-Hochberg control across the four endpoints. Year-to-year variation in the Sunday contrast is small relative to its sampling error for mental health, so that test is uninformative; attenuation-corrected (Deming) slopes were unstable for the same reason and are not reported. A pandas recomputation matched the slopes to less than 1e-14.

Holiday deviation against the period's Sunday contrast

Endpoint	Sample	Mean Sunday contrast (SD)	Slope (95% CI)	r	Benjamini-Hochberg q
Drug or poisoning	All 24	0.190 (0.096)	0.71 (0.33 to 1.08)	0.42	0.00304
Drug or poisoning	Excluding 2020 (22)	0.207 (0.071)	0.62 (-0.16 to 1.40)	0.27	
Assault	All 24	0.415 (0.209)	0.06 (-0.45 to 0.57)	0.04	0.883
Assault	Excluding 2020 (22)	0.408 (0.217)	0.15 (-0.36 to 0.67)	0.10	
Mental health	All 24	-0.189 (0.067)	0.05 (-0.61 to 0.70)	0.03	0.883
Mental health	Excluding 2020 (22)	-0.193 (0.068)	0.04 (-0.66 to 0.74)	0.03	
All calls	All 24	-0.060 (0.029)	1.02 (-0.03 to 2.07)	0.39	0.112
All calls	Excluding 2020 (22)	-0.060 (0.029)	1.08 (-0.03 to 2.20)	0.41	

Supplementary Table S37. Demand and access: holiday response times against the Sunday pattern, and marathon held calls. Both designs were fixed and hashed before fitting. Panel A (14 October) compares each holiday with the same weekday 7 and 14 days before and after and with every ordinary Sunday 4-17 days away, in the same models for call counts (city-hour conditional Poisson with temperature and precipitation) and for ZIP-hour mean call-to-scene time (weighted by calls, hour-of-day terms, as in Supplementary Table S25). The Sunday contrast is the day-off part; the holiday-beyond-Sunday residual is what remains. Memorial and Labor Day pool 24 Mondays; July 4 uses the 9 weekday occurrences. Intervals are cluster-robust with t references (clusters minus one degrees of freedom). Gates reproduced the published holiday response estimates exactly (-67.52 and -89.43 seconds). Panel B (1 October) uses the principal marathon candidate sample and design; reproduction gates matched the headline active-window estimates exactly (151.84

seconds equal-weight; 151.10 seconds incident-level). Held calls use the public HELD_INDICATOR field (true/false in 2013-2018, Y/N in 2019-2023; base rates show no break). Dispatch and travel seconds partition the same principal outcome and sample; they differ from the separate component models in Supplementary Fig. S1, which use their own validity filters.

Panel A: holiday call-to-scene time and call volume, day-off part and holiday-beyond-Sunday residual

Holiday	Quantity	Call-to-scene time	All calls
Memorial and Labor Day (24)	Holiday vs same weekday	-97.0 s (-125.7 to -68.3)	-6.9% (-9.0 to -4.7)
Memorial and Labor Day (24)	Ordinary Sunday vs same weekday	-57.3 s (-89.3 to -25.4)	-6.2% (-7.1 to -5.4)
Memorial and Labor Day (24)	Holiday beyond Sunday	-39.6 s (-63.8 to -15.5)	-0.7% (-2.5 to 1.1)
Weekday July 4 (9)	Holiday vs same weekday	-90.8 s (-119.4 to -62.1)	-6.9% (-8.8 to -4.9)
Weekday July 4 (9)	Ordinary Sunday vs same weekday	-31.3 s (-61.2 to -1.3)	-3.6% (-5.3 to -2.0)
Weekday July 4 (9)	Holiday beyond Sunday	-59.5 s (-94.8 to -24.2)	-3.4% (-5.2 to -1.5)

Panel B: marathon held calls and never-held access, route-adjacent vs distant, equal-weight over 10 races, t(9)

Outcome	Window	Estimate (95% CI)	Races positive
Call-to-scene time, all calls (s)	Active operations	+151.8 (93.4 to 210.3)	10 of 10
Calls held for unit assignment (percentage points)	Active operations	+5.2 (0.6 to 9.9)	8 of 10
Calls held for unit assignment (percentage points)	Prior Saturday	-0.1 (-7.8 to 7.6)	5 of 10
Calls held for unit assignment (percentage points)	Active minus prior Saturday	+5.3 (-3.7 to 14.3)	6 of 10
Call-to-scene time, never-held calls (s)	Active operations	+114.2 (78.9 to 149.6)	10 of 10
Call-to-scene time, never-held calls (s)	Prior Saturday	-2.6 (-69.6 to 64.4)	5 of 10
Call-to-scene time, never-held calls (s)	Active minus prior Saturday	+116.9 (52.6 to 181.1)	9 of 10
Dispatch seconds, principal sample	Active operations	+68.4 (7.9 to 129.0)	9 of 10
Dispatch seconds, principal sample	Prior Saturday	+3.0 (-55.3 to 61.3)	3 of 10
Dispatch seconds, principal sample	Active minus prior Saturday	+65.5 (-14.2 to 145.1)	8 of 10
Travel seconds, principal sample	Active operations	+83.4 (53.5 to 113.3)	10 of 10
Travel seconds, principal sample	Prior Saturday	+0.8 (-64.9 to 66.5)	6 of 10
Travel seconds, principal sample	Active minus prior Saturday	+82.6 (16.1 to 149.1)	9 of 10

Panel B (continued): active-window held rates

Day	Area	Calls	Held (%)
Matched Sundays	Route-adjacent	12,123	3.0
Marathon Sunday	Route-adjacent	3,816	11.2
Matched Sundays	Distant	36,296	2.6
Marathon Sunday	Distant	9,199	4.6

Supplementary Table S38. Population presence: foot traffic around the calendar pulses and presence-adjusted call estimates. The design was fixed and hashed before fitting. Presence is measured from Advan Weekly Patterns Plus visits to commercial points of interest (POIs) in the 170 panel ZIPs, 7 January 2019 to 31 December 2023 (licensed through Dewey; not redistributed), as visits per POI (primary) or total visits. July 2022 is excluded because the POI panel dipped in three of its weeks; sensitivities include it or drop 2020-2021, as in the paper's pandemic exclusion. Panel A applies the early-month matched design to log presence. Panel B fits the joint calendar specification of Supplementary Table S31 to log presence on 2019-2023; quantities are defined as in Supplementary Table S37. Panel C compares early-month and holiday call estimates on the same 2019-2023 sample without and with log presence; the retained share is the adjusted log estimate divided by the unadjusted one. Because presence lies on the day-off pathway, holiday adjustment describes departures from what presence predicts rather than confounding. Panel D repeats the early-month presence contrast for each ZIP, meta-regresses it on the moderators of Supplementary Table S33 (univariate REML), and refits the early-month call contrast on ZIP-days with ZIP presence. Gates: the published early-month IRR was reproduced exactly on 2012-2023; every included date has all 170 ZIPs; holidays are local minima of visits per POI in 14 of 15 year-holidays; all joint-model terms are identified. Independent statsmodels and base-R refits match the reported point estimates. Commercial foot traffic is an incomplete measure of residential population.

Panel A: foot traffic on days 1-3 of the month

Sample	Presence	Case days	Days 1-3 vs matched days
Primary (July 2022 excluded)	Visits per POI	162	-1.4% (-3.8 to 1.0)
Primary (July 2022 excluded)	Total visits	162	-1.7% (-4.1 to 0.8)
Including July 2022	Visits per POI	165	-1.5% (-3.8 to 0.8)
Including July 2022	Total visits	165	-1.8% (-4.1 to 0.7)
Excluding 2020-2021	Visits per POI	96	-1.3% (-3.1 to 0.5)
Excluding 2020-2021	Total visits	96	-1.3% (-3.1 to 0.5)

Panel B: foot traffic on holidays against the same weekday and the Sunday pattern (primary sample)

Holiday	Quantity	Visits per POI
Memorial Day	Holiday vs same weekday	-12.2% (-21.7 to -1.6)
Memorial Day	Ordinary Sunday vs same weekday	-9.7% (-15.4 to -3.7)
Memorial Day	Holiday beyond Sunday	-2.8% (-8.7 to 3.5)
Labor Day	Holiday vs same weekday	-22.1% (-28.2 to -15.5)
Labor Day	Ordinary Sunday vs same weekday	-13.1% (-17.4 to -8.5)
Labor Day	Holiday beyond Sunday	-10.4% (-16.7 to -3.6)
Memorial and Labor Day	Holiday vs same weekday	-17.3% (-22.4 to -11.9)
Memorial and Labor Day	Ordinary Sunday vs same weekday	-11.4% (-15.0 to -7.6)
Memorial and Labor Day	Holiday beyond Sunday	-6.6% (-10.2 to -2.9)
Weekday July 4	Holiday vs same weekday	-25.4% (-37.0 to -11.6)
Weekday July 4	Ordinary Sunday vs same weekday	-19.2% (-24.7 to -13.4)
Weekday July 4	Holiday beyond Sunday	-7.6% (-18.2 to 4.3)

Panel C1: early-month case-crossover on 2019-2023, without and with log visits per POI

Sample	Endpoint	Unadjusted	Adjusted for presence
Primary (July 2022 excluded)	All calls	+1.6% (0.5 to 2.9)	+1.7% (0.4 to 2.9)
Primary (July 2022 excluded)	Drug or poisoning	+6.9% (5.1 to 8.8)	+7.3% (5.5 to 9.1)
Including July 2022	All calls	+1.7% (0.5 to 2.8)	+1.7% (0.5 to 2.9)
Including July 2022	Drug or poisoning	+6.8% (5.0 to 8.6)	+7.2% (5.5 to 9.0)
Excluding 2020-2021	All calls	+1.2% (0.3 to 2.1)	+1.5% (0.7 to 2.3)
Excluding 2020-2021	Drug or poisoning	+7.4% (5.1 to 9.8)	+7.6% (5.2 to 9.9)

Panel C2: joint calendar model on 2019-2023, without and with log visits per POI (primary sample)

Endpoint	Pulse	Quantity	Unadjusted	Adjusted for presence	Retained share
Drug or poisoning	Days 1-3 (mean)	Days 1-3 vs days 10-24	+6.7%	+7.3% (5.2 to 9.4)	1.08
Drug or poisoning	Memorial and Labor Day	Holiday vs same weekday	+11.6%	+18.0% (7.8 to 29.1)	1.51
Drug or poisoning	Memorial and Labor Day	Holiday beyond Sunday	-5.2%	-3.2% (-8.1 to 1.9)	0.61
Drug or poisoning	Weekday July 4	Holiday vs same weekday	+24.7%	+35.8% (15.3 to 59.8)	1.39
Drug or poisoning	Weekday July 4	Holiday beyond Sunday	+8.4%	+11.1% (-3.7 to 28.3)	1.30
Assault	Days 1-3 (mean)	Days 1-3 vs days 10-24	+2.2%	+2.4% (-3.3 to 8.4)	1.08
Assault	Memorial and Labor Day	Holiday vs same weekday	+42.8%	+45.5% (23.7 to 71.0)	1.05
Assault	Memorial and Labor Day	Holiday beyond Sunday	+3.2%	+4.0% (-14.9 to 26.9)	1.22
Assault	Weekday July 4	Holiday vs same weekday	+62.8%	+67.2% (54.9 to 80.5)	1.06
Assault	Weekday July 4	Holiday beyond Sunday	+27.2%	+28.1% (11.4 to 47.2)	1.03

Panel D: ZIP level (170 ZIPs; 82,045 matched ZIP-days). The corresponding call slope is +3.4% per SD of Black share (Supplementary Table S33).

Quantity	Estimate
Pooled ZIP presence change, days 1-3	-0.9% (-1.1 to -0.7)
Per SD log median income	-0.5% (-0.8 to -0.2)
Per SD non-Hispanic Black share	+0.3% (0.1 to 0.5)
Per SD Hispanic share	-0.2% (-0.5 to 0.1)
ZIP-day drug or poisoning calls, days 1-3, unadjusted	+6.9% (4.6 to 9.2)
ZIP-day drug or poisoning calls, days 1-3, adjusted for ZIP presence	+7.0% (5.0 to 9.1)

Supplementary Table S39. Calibration of 2024-2025 calendar effects against 2012-2023. No-intercept slopes of the 2024-2025 log incidence rate ratio on the 2012-2023 log incidence rate ratio, for the seven locked direction cells and for all 15 schedule-by-endpoint cells of Supplementary Table S18, with inverse-variance weights from the 2024-2025 intervals (primary) and without weights; t intervals use cells minus one degrees of freedom. The retained share is each locked cell's 2024-2025 log effect divided by its 2012-2023 log effect. Training effects carry sampling error, so a slope below 1 partly reflects regression to the mean. This summary was specified after the directional validation was known.

Panel A: calibration slopes

Cells	Weights	Slope (95% CI)
Seven locked cells	Inverse variance	0.61 (0.47 to 0.75)
Seven locked cells	Unweighted	0.56 (0.45 to 0.67)
All 15 cells	Inverse variance	0.62 (0.52 to 0.72)
All 15 cells	Unweighted	0.57 (0.45 to 0.69)

Panel B: locked cells

Schedule	Endpoint	IRR 2012-2023	IRR 2024-2025	Retained share
Early month	Drug or poisoning	1.086	1.090	1.05
July 4	Assault	2.373	1.786	0.67
July 4	Burns	3.391	1.905	0.53
July 4	Drug or poisoning	1.350	1.120	0.38
Memorial/Labor Day	Assault	1.440	1.119	0.31
Memorial/Labor Day	Mental health	0.845	0.838	1.05
Memorial/Labor Day	Drug or poisoning	1.181	1.129	0.73

Supplementary Table S40. Labor Day parade check for the holiday racial gradient. J'Ouvert and the West Indian American Day Parade take place on Labor Day along Eastern Parkway in Central Brooklyn. Before fitting, ZIPs were selected whose ZCTA centroids lie within 1.5 km of Eastern Parkway between Grand Army Plaza and Utica Avenue (11213, 11238, 11225, 11216, 11217, 11233). The two-stage multivariate meta-regression of Supplementary Table S33 (median income, non-Hispanic Black and Hispanic shares) was refitted without them, keeping full-sample standardization; the full-sample fit reproduced the published slopes exactly. The prespecified reading attributes the Memorial/Labor gradient in part to the parade neighbourhoods if its slope falls by more than half or its interval includes zero.

Black-share slope with and without the parade ZIPs

Pulse	Sample	ZIPs	Per SD of Black share (95% CI)
Memorial and Labor Day	all ZIPs	170	+7.8% (2.3 to 13.5)
Memorial and Labor Day	excluding parade ZIPs	164	+4.0% (-0.7 to 8.9)
Early month (days 1-3)	all ZIPs	170	+3.4% (2.3 to 4.5)
Early month (days 1-3)	excluding parade ZIPs	164	+3.1% (2.0 to 4.3)

Supplementary Table S41. Absolute scale of the calendar and street-event pulses. Assembled from existing estimates without new models. Early-month excess applies the published incidence rate ratio to the mean of the matched control days (February-December 2012-2023) for three case days per month and 11 months per year. Holiday excess is the unadjusted treated-minus-control rate multiplied by the hourly observations per event window, whereas relative changes are the adjusted incidence rate ratios, so the two need not reconcile exactly. Marathon long-wait counts are the descriptive translation of Supplementary Table S19. The street-event row counts all classified exposed precinct-hours as occurrences (mean per year, 2024-2025); the analytic panels of Supplementary Tables S16 and S49 retain the 3,706 exposed precinct-hours across both years that have observed call-to-scene data. Excess calls describe dispatcher-coded counts, not attributable cases.

Baselines, relative changes and absolute excess

Pulse	Occurrences per year	Baseline	Baseline unit	Relative change (%)	Excess per occurrence	Unit	Excess per year
Early month (days 1-3), drug or poisoning	11.0	253.5	calls per control day	8.6	65.4	calls per month (3 days)	719.9
Early month (days 1-3), all calls (context)	11.0	3,770.2	calls per control day				
July 4 fireworks window (18:00-05:59), Drug or poisoning	1.0	157.6	expected calls per occurrence (matched windows)	34.1	55.2	calls per occurrence	55.2
July 4 fireworks window (18:00-05:59), Assault	1.0	20.7	expected calls per occurrence (matched windows)	132.9	28.4	calls per occurrence	28.4
July 4 fireworks window (18:00-05:59), Mental Health	1.0	171.2	expected calls per occurrence (matched windows)	-8.7	-14.4	calls per occurrence	-14.4
July 4 fireworks window (18:00-05:59), Cardiac Arrest	1.0	29.8	expected calls per occurrence (matched windows)	5.8	2.1	calls per occurrence	2.1
July 4 fireworks window (18:00-05:59), Burns	1.0	4.1	expected calls per occurrence (matched windows)	233.5	9.8	calls per occurrence	9.8
Memorial or Labor Day (full day), Drug or poisoning	2.0	234.8	expected calls per occurrence (matched windows)	16.4	42.6	calls per occurrence	85.2
Memorial or Labor Day (full day), Assault	2.0	25.4	expected calls per occurrence (matched windows)	39.3	11.2	calls per occurrence	22.4
Memorial or Labor Day (full day), Mental Health	2.0	409.6	expected calls per occurrence (matched windows)	-15.8	-63.3	calls per occurrence	-126.7

Pulse	Occurrences per year	Baseline	Baseline unit	Relative change (%)	Excess per occurrence	Unit	Excess per year
Memorial or Labor Day (full day), Cardiac Arrest	2.0	70.2	expected calls per occurrence (matched windows)	-0.1	0.4	calls per occurrence	0.8
Memorial or Labor Day (full day), Burns	2.0	7.2	expected calls per occurrence (matched windows)	3.1	0.4	calls per occurrence	0.9
Marathon, route-adjacent call-to-scene (active window)	1.0	476.2	mean seconds, route-adjacent control Sundays	31.9	151.8	seconds per call	
Marathon, route-adjacent waits over 15 minutes	1.0	8.8	% of route-adjacent incidents (control)	77.1	36.7	incidents per race (descriptive)	36.7
UN General Assembly week, UN-adjacent call-to-scene	1.0				112.6	seconds per call (equal-year)	
Multi-precinct full-street event, precinct-hour call-to-scene (2024-2025)	2,986.5				37.4	seconds per call; occurrences = exposed precinct-hours per year	

Supplementary Table S42. Cincinnati Flying Pig Marathon access check. An August 2026 design, frozen before outcomes, ported the marathon contrast to Cincinnati Fire/EMS dispatch records (create-to-arrival time; incidents within 1 km of official City closure geometry against those beyond 5 km; race Sunday against matched control Sundays; active window 05:00-14:59) for the four 2022-2025 races with published closure notices. Panel A reports that result. Route-adjacent race-day records lost substantially more valid intervals than other cells, mostly through missing arrival times, so the estimate describes calls with a recorded arrival. Panel B applies the same design with the course reconstructed along its documented streets, which places the near band on the full-course corridor. Panel C reports the prespecified selection gate for the six earlier in-person races (2015-2019 and 31 October 2021), evaluated on validity alone before any response time was computed: the route-adjacent minus distant difference in invalid shares exceeded the 10-percentage-point stop rule, so those races were not fitted. The same rule would also have stopped the 2022-2025 races. Intervals use t references with 3 degrees of freedom.

Panel A: 2022-2025 races, official closure geometry

Quantity	Window	Estimate (95% CI)	Races positive
Create-to-arrival (primary)	Active (05-14 h)	+67.34 (-17.36 to 152.04)	4 of 4
Create-to-dispatch	Active (05-14 h)	+38.20 (-48.38 to 124.77)	3 of 4
Dispatch-to-arrival (travel)	Active (05-14 h)	+62.19 (33.79 to 90.59)	4 of 4
Valid-interval share (proportion)	Active (05-14 h)	-0.19 (-0.41 to 0.03)	1 of 4
Setup or post window	Setup (01-04 h)	+6.29 (-56.52 to 69.10)	2 of 4
Setup or post window	Post (15-18 h)	-3.39 (-127.23 to 120.44)	1 of 4

Panel B: 2022-2025 races, reconstructed course

Quantity	Estimate (95% CI)	Races positive
Create-to-arrival (primary)	+81.18 (-13.77 to 176.14)	4 of 4
Create-to-dispatch	+42.20 (-48.39 to 132.79)	3 of 4
Dispatch-to-arrival (travel)	+77.40 (47.28 to 107.52)	4 of 4
Valid-interval share (proportion)	-0.22 (-0.41 to -0.03)	0 of 4

Panel C: selection gate for the six earlier races (stop if the pooled near-far difference exceeds 10 percentage points)

Race	Route-adjacent race-day incidents	Invalid share, race-day near (%)	Invalid share, control near (%)	Near-far difference in invalid share (pp)
Pooled six races	160	25.0	16.0	+11.2
2015	31	16.1	18.3	-14.2
2016	25	36.0	11.9	+29.6
2017	18	33.3	17.6	+22.3
2018	28	35.7	18.2	+16.7
2019	33	21.2	18.6	+11.4
2021	25	12.0	11.2	+1.9

Supplementary Table S43. How much daily EMS load the calendar anticipates. The design was fixed and hashed before fitting. Nested Poisson models of citywide daily counts were fitted on 2012-2023 and applied without refitting to 2024-2025: M0, weekday-by-calendar-month rhythm; M1, M0 plus the holiday lead-lag and day-of-month terms of the joint calendar model, knowable months ahead; M2, M1 plus realized Central Park temperature (lags 0-3) and same-day precipitation, an upper bound for days-ahead information. Each month's observed total is given to every model, so the metrics score the within-month shape of daily load rather than level forecasting. Panel A reports the share of within-month day-to-day deviance explained in 2024-2025 and the gains from calendar pulse terms (M1 minus M0) and weather (M2 minus M1); Panel B reports the share of the highest-load tenth of days (relative to their month) that each model ranks among its own top tenth; intervals use a 24-month block bootstrap. Panel C decomposes 2012-2023 within-month deviance (Shapley shares). The joint-model gate reproduced all published coefficients; an independent Python implementation reproduced every metric. 2024-2025 counts were partly seen in earlier analyses, so this is a retrospective out-of-sample evaluation. Post hoc, most of the total-call gain came from lower-load federal holidays and most of the drug or poisoning gain from the New Year window.

Panel A: percent of within-month day-to-day deviance explained in 2024-2025 (percentage points for gains)

Quantity	All calls	Drug or poisoning
Weekday-by-month rhythm (M0)	28.8 (22.9 to 35.6)	32.9 (24.7 to 40.6)
Plus calendar pulses (M1)	35.9 (27.4 to 44.2)	44.7 (35.9 to 52.5)
Plus realized weather (M2)	63.8 (53.2 to 71.8)	56.0 (50.0 to 61.2)
Gain from calendar pulses (M1 - M0)	7.1 (-2.1 to 14.0)	11.9 (3.4 to 21.8)
Gain from weather (M2 - M1)	27.9 (19.1 to 37.3)	11.2 (6.3 to 17.4)

Panel B: percent of the highest-load tenth of days flagged

Quantity	All calls	Drug or poisoning
Chance	10.0 (9.9 to 10.1)	10.0 (9.9 to 10.1)
M0	18.5 (11.5 to 26.4)	39.5 (31.5 to 48.3)
M1	23.3 (16.2 to 34.2)	47.9 (38.4 to 56.2)
M2	53.4 (44.6 to 64.4)	52.1 (45.2 to 62.2)
M1 minus M0	4.8 (-1.8 to 15.6)	8.5 (1.9 to 13.7)

Panel C: 2012-2023 within-month deviance shares

Component	All calls	Drug or poisoning
Calendar pulses	11.4%	18.9%
Residual	42.0%	31.0%
Weather	23.7%	7.1%
Weekday-by-month rhythm	22.9%	43.0%

Supplementary Table S44. Call-type transitions, dispositions and the coding pathway for the early-month increase. The design was fixed and hashed before fitting. All quantities use the code-75 early-month case-crossover (days 1-3 against same-weekday days 10-24 in the same month-year, February-December 2012-2023, temperature-adjusted, clustered by stratum). Shares are modelled as counts with a log-denominator offset, so estimates are ratios of proportions. Panel A tracks incidents between initial and final dispatcher classes: the forward transition is the share of unconscious, cardiac-arrest or unknown intakes closed as drug or poisoning; the reverse is the share of drug or poisoning intakes closed in those classes; never-drug series exclude incidents coded drug or poisoning at intake. Panel B reports dispositions among final drug or poisoning calls. Panel C refits the two-stage ZIP meta-regression of Supplementary Table S33 (income, Black and Hispanic shares) for each endpoint. Gates reproduced the published city-hour counts in all clock cells, the published IRRs, the ZIP-day panel and the published Black-share slope exactly; an independent Python implementation and metafor refit matched all estimates. Dispatch codes are not clinical diagnoses, and the data carry no naloxone field.

Panel A: transitions between initial and final call types, days 1-3 against matched days

Quantity	Measure	Estimate (95% CI)	Primary
P(final drug | initial unconscious/arrest/unknown)	ratio of proportions	1.055 (1.014 to 1.098)	yes
Count: initial unconscious/arrest/unknown -> final drug	IRR	1.095 (1.051 to 1.141)	
P(final drug | initial unconscious)	ratio of proportions	1.044 (0.982 to 1.110)	
P(final drug | initial cardiac arrest)	ratio of proportions	0.948 (0.825 to 1.089)	
P(final drug | initial unknown)	ratio of proportions	1.077 (1.021 to 1.137)	
P(final unconscious/arrest/unknown | initial drug)	ratio of proportions	0.989 (0.933 to 1.048)	yes
Count: initial drug -> final unconscious/arrest/unknown	IRR	1.079 (1.016 to 1.146)	
Final unconscious, all origins	IRR	1.062 (1.047 to 1.077)	
Final unconscious, initial not drug	IRR	1.061 (1.046 to 1.076)	
Final cardiac arrest, all origins	IRR	1.045 (1.024 to 1.066)	
Final cardiac arrest, initial not drug	IRR	1.045 (1.024 to 1.067)	
Final seizure, all origins	IRR	1.021 (1.007 to 1.035)	
Final seizure, initial not drug	IRR	1.021 (1.007 to 1.035)	
Final drug, all origins (published endpoint)	IRR	1.086 (1.074 to 1.098)	
Final drug, initial drug	IRR	1.088 (1.076 to 1.101)	
Final drug, initial unconscious/arrest/unknown	IRR	1.095 (1.051 to 1.141)	
Final drug, initial mental health (EDP/altered)	IRR	1.079 (1.039 to 1.121)	
Final drug, initial other non-drug	IRR	1.038 (0.998 to 1.080)	
Initial drug (any final type)	IRR	1.089 (1.077 to 1.101)	

Quantity	Measure	Estimate (95% CI)	Primary
Final drug or unconscious (combined)	IRR	1.075 (1.065 to 1.086)	

Panel B: dispositions

Quantity	Measure	Estimate (95% CI)	Primary
Refused medical aid (93) | final drug	ratio of proportions	1.016 (0.994 to 1.038)	yes
Transporting patient (82) | final drug	ratio of proportions	1.003 (0.997 to 1.009)	
Patient gone on arrival (96) | final drug	ratio of proportions	0.998 (0.972 to 1.025)	
Unfounded (90) | final drug	ratio of proportions	0.976 (0.956 to 0.997)	
Refused (93) | final drug with presumptive patient contact	ratio of proportions	1.011 (0.990 to 1.033)	
Code-77 non-transport bucket {90,93,96} | final drug	ratio of proportions	0.996 (0.984 to 1.008)	
Refused medical aid (93) | final not drug	ratio of proportions	1.009 (1.000 to 1.018)	
Refusal RP, drug / non-drug (stacked)	ratio of ratios	1.006 (0.984 to 1.029)	

Panel C: early-month Black-share slope by endpoint

Endpoint	ZIPs	Per SD of Black share (95% CI)
Drug or poisoning (final; published)	169	+3.44% (2.35 to 4.54)
Unconscious	169	+3.18% (1.95 to 4.42)
Drug or poisoning plus unconscious	169	+3.23% (2.30 to 4.16)
Drug or poisoning at intake	169	+3.33% (2.19 to 4.49)

Supplementary Table S45. Street-event delays under concurrent load and on holiday weekends, and where the added waiting fell. The design was fixed and hashed before fitting and reuses the 2024-2025 permit-census classification, panel, model and inference (Supplementary Table S16); the published multi-precinct estimate was reproduced to 1e-13. Expected citywide demand for each hour comes from a calendar model fitted only on hours without any multi-precinct event (realized load is affected by events and was not used); event hours are split into thirds. Holiday weekends cover 35 dates around federal holidays. Long-wait outcomes use precinct-hour aggregates pulled from NYC Open Data (76xm-jjuj), aligned with the frozen response data on 99.9998% of rows. Panel A reports the moderated contrasts; a federal-holiday-only sensitivity (13 dates, mostly flagship parades) gave a much larger difference that reflects event identity and scale rather than load. Panel B distributes the pooled excess call-seconds over exposed precinct-hours (a descriptive allocation, not precinct-specific effects) and summarizes it by borough and by thirds of precinct poverty and nonwhite share. Low-load event hours are mostly 07:00-12:00, so load cannot be separated from time of day or event phase. An independent Python implementation matched all coefficients and contrasts.

Panel A: moderated street-event contrasts

Contrast	Outcome	Estimate (95% CI)	P
Highest minus lowest third of expected load	Call-to-scene (s)	-26.6 (-76.2 to 23.0)	0.296
Middle minus lowest third	Call-to-scene (s)	+0.9 (-38.7 to 40.4)	0.966
Highest minus middle third	Call-to-scene (s)	-27.4 (-73.1 to 18.2)	0.242
Holiday weekends minus other dates	Call-to-scene (s)	+45.0 (-9.3 to 99.2)	0.108
Federal holiday dates minus other dates (sensitivity)	Call-to-scene (s)	+149.2 (121.5 to 177.0)	0.000
Highest minus lowest third of expected load	Waits over 10 min (pp)	-4.0 (-7.2 to -0.9)	0.015
Middle minus lowest third	Waits over 10 min (pp)	-2.7 (-5.9 to 0.5)	0.098
Highest minus middle third	Waits over 10 min (pp)	-1.3 (-3.8 to 1.2)	0.310
Holiday weekends minus other dates	Waits over 10 min (pp)	+3.8 (-0.6 to 8.2)	0.096
Highest minus lowest third of expected load	Waits over 15 min (pp)	-1.7 (-4.4 to 1.0)	0.229
Middle minus lowest third	Waits over 15 min (pp)	-0.5 (-3.3 to 2.3)	0.715
Highest minus middle third	Waits over 15 min (pp)	-1.2 (-4.2 to 1.8)	0.448
Holiday weekends minus other dates	Waits over 15 min (pp)	+3.1 (-0.2 to 6.4)	0.072

Panel B: where the added waiting fell, 2024-2025

Group	Share of excess waiting	Share of residents
MANHATTAN	53%	20%
BRONX	13%	18%
BROOKLYN	20%	32%
QUEENS	9%	24%
RICHMOND / STATEN ISLAND	4%	6%
Poverty third 1 (5-13%)	42%	34%
Poverty third 2 (13-19%)	25%	33%
Poverty third 3 (19-39%)	33%	33%

Group	Share of excess waiting	Share of residents
Nonwhite share third 1 (9-41%)	45%	33%
Nonwhite share third 2 (42-69%)	24%	34%
Nonwhite share third 3 (70-94%)	30%	33%

Supplementary Table S46. Calls and response times combined as pre-arrival waiting, by schedule. The design was fixed and hashed on 21 October 2026, after the paper's separate count and response-time results were known, so it is an exploratory synthesis of existing designs. Each schedule keeps its reported comparison and geography: early-month days 1-3 against same-weekday days 10-24 (citywide, 170 ZIPs; 396 case days); the July 4 fireworks window (18:00-05:59) and the Memorial and Labor Day holiday day against the same hours 7 and 14 days before and after (citywide); Memorial and Labor Day also against the Sundays preceding those control Mondays; marathon active operations (07:00-20:59) in route-adjacent ZIPs, with distant ZIPs and four control Sundays forming the counterfactual; and United Nations General Assembly weekdays in UN-adjacent ZIPs, with other Manhattan ZIPs and the same weekdays 2 and 4 weeks away. Total waiting is the number of incidents multiplied by their mean valid call-to-scene time (dispatch plus travel, 0-3,600 s; invalid intervals take the cell mean). Its excess over the counterfactual splits exactly into a part from more calls at the usual wait, a part from the usual calls at the changed wait, and their interaction; the split is an accounting, not an attribution. Relative change divides the excess by mean comparison waiting; because totals scale with each area's baseline volume, relative changes, not incident-hours, compare across rows. Intervals are percentile intervals from 2,000 cluster bootstrap draws over month-years, occurrences, races or years; marathon (10 races) and UN (11 years) intervals are approximate and narrower than t intervals for the same quantities (marathon wait 108-202 against the published 93-210 seconds). Shares are shown only where the total excludes zero. Gates reproduced every published hourly call count for the 170 ZIPs, the early-month IRR (1.086073), the holiday matched-window counts, and all 10 published marathon race effects to 1e-6; the UN equal-year estimate under the published 7,200-s ceiling was 105.5 s against 112.6 s published (bridge tolerance 15 s). An independent R implementation reproduced all 78 primary point estimates (largest difference 6e-14). The July 4 full holiday day, estimated after the first run to separate window from outcome, gave -64 s, close to the published full-day -67.5 s. Panel B excludes standby call types (the published volume definition) or the source standby flag. Placebo rows are single occurrences and are shown without intervals.

Panel A: primary estimates

Schedule	Calls, %	Wait, s	Validity, pp	Excess waiting, patient-h	Relative, %	Added calls without valid interval	More calls / slower / both, h	Shares, calls / slower	Per year, patient-h
Early month (per day)	+1.2 (0.2 to 2.4)	+2.6 (-4.7 to 10.2)	-0.1	+10.6 (-2.0 to 25.0)	+1.8 (-0.3 to 4.4)	7%	6.2 / 1.6 / 2.8		+349
July 4 fireworks window	+19.9 (15.4 to 24.2)	+87.6 (49.2 to 124.0)	-2.4	+104.1 (71.0 to 136.5)	+40.9 (27.6 to 53.8)	16%	50.5 / 43.1 / 10.5	48% / 41%	+104

Schedule	Calls, %	Wait, s	Validity, pp	Excess waiting, patient-h	Relative, %	Added calls without valid interval	More calls / slower / both, h	Shares, calls / slower	Per year, patient-h
Memorial/Labor Day vs matched days	-5.8 (-8.3 to -3.1)	-80.3 (-99.1 to -61.0)	+0.0	-116.6 (-147.8 to -83.8)	-18.8 (-23.1 to -13.7)	2%	-35.3 / -88.4 / 7.1	30% / 76%	-233
Memorial/Labor Day vs Sunday	+0.4 (-2.1 to 2.8)	-35.5 (-49.2 to -21.4)	-0.1	-33.1 (-56.5 to -7.7)	-6.2 (-10.4 to -1.4)	29%	3.1 / -37.3 / 1.0	-10% / 113%	-66
Marathon, route-adjacent (per race)	+24.2 (15.1 to 35.2)	+151.8 (107.9 to 202.3)	-4.9	+28.3 (17.5 to 43.5)	+64.1 (41.4 to 91.3)	30%	11.1 / 13.2 / 4.1	39% / 46%	+28
UN week, UN-adjacent (per day)	+6.2 (2.3 to 10.4)	+68.5 (43.0 to 95.9)	-2.4	+2.4 (1.6 to 3.2)	+19.0 (13.0 to 25.2)	44%	0.8 / 1.5 / 0.1	34% / 62%	+12

Panel B: standby exclusions

Schedule	Calls, %	Wait, s	Validity, pp	Excess waiting, patient-h	Relative, %	Added calls without valid interval	More calls / slower / both, h	Shares, calls / slower	Per year, patient-h	Exclusion
Early month (per day)	+1.2 (0.2 to 2.3)	+2.6 (-4.9 to 10.2)	-0.1	+10.5 (-2.7 to 24.8)	+1.8 (-0.5 to 4.3)	7%	6.1 / 1.6 / 2.8		+348	standby call types
July 4 fireworks window	+19.8 (15.2 to 24.3)	+87.6 (49.3 to 124.9)	-2.4	+103.5 (70.4 to 136.1)	+40.7 (27.2 to 54.3)	16%	50.1 / 43.0 / 10.4	48% / 42%	+104	standby call types
Memorial/Labor Day vs matched days	-5.8 (-8.5 to -3.1)	-80.3 (-99.6 to -61.9)	+0.0	-116.7 (-149.0 to -83.8)	-18.9 (-23.4 to -14.1)	2%	-35.6 / -88.2 / 7.1	31% / 76%	-233	standby call types
Memorial/Labor Day vs Sunday	+0.4 (-2.1 to 2.8)	-35.4 (-49.4 to -20.9)	-0.1	-33.1 (-56.6 to -7.9)	-6.2 (-10.4 to -1.5)	29%	3.0 / -37.1 / 1.0	-9% / 112%	-66	standby call types
Marathon, route-adjacent (per race)	+24.4 (15.5 to 35.8)	+152.3 (108.3 to 203.9)	-4.9	+28.4 (17.7 to 43.9)	+64.4 (42.3 to 91.5)	30%	11.1 / 13.2 / 4.1	39% / 46%	+28	standby call types
UN week, UN-adjacent (per day)	+4.6 (0.6 to 8.3)	+71.0 (47.1 to 99.3)	-1.5	+2.2 (1.4 to 3.1)	+17.7 (11.6 to 23.7)	37%	0.6 / 1.5 / 0.1	27% / 70%	+11	standby call types
Early month (per day)	+1.2 (0.2 to 2.4)	+2.6 (-4.9 to 10.2)	-0.1	+10.5 (-2.9 to 25.1)	+1.8 (-0.5 to 4.4)	7%	6.1 / 1.5 / 2.8		+347	source standby flag
July 4 fireworks window	+19.8 (15.3 to 24.2)	+87.5 (48.3 to 122.0)	-2.4	+103.6 (70.1 to 135.3)	+40.7 (27.1 to 53.6)	16%	50.2 / 43.0 / 10.4	48% / 41%	+104	source standby flag

Schedule	Calls, %	Wait, s	Validity, pp	Excess waiting, patient-h	Relative, %	Added calls without valid interval	More calls / slower / both, h	Shares, calls / slower	Per year, patient-h	Exclusion
Memorial/Labor Day vs matched days	-5.8 (-8.3 to -3.1)	-80.3 (-99.8 to -61.9)	+0.0	-116.8 (-148.9 to -83.2)	-18.9 (-23.2 to -13.8)	2%	-35.6 / -88.4 / 7.1	30% / 76%	-234	source standby flag
Memorial/Labor Day vs Sunday	+0.3 (-2.1 to 2.9)	-35.5 (-49.5 to -22.5)	-0.1	-33.2 (-58.1 to -8.6)	-6.2 (-10.7 to -1.6)	29%	3.0 / -37.2 / 1.0	-9% / 112%	-66	source standby flag
Marathon, route-adjacent (per race)	+24.3 (14.7 to 35.5)	+151.9 (107.3 to 206.0)	-4.9	+28.4 (17.4 to 44.1)	+64.2 (41.0 to 91.5)	30%	11.1 / 13.1 / 4.1	39% / 46%	+28	source standby flag
UN week, UN-adjacent (per day)	+5.4 (1.0 to 9.4)	+68.3 (44.1 to 98.7)	-1.8	+2.3 (1.5 to 3.1)	+18.0 (11.9 to 24.2)	39%	0.7 / 1.5 / 0.1	31% / 65%	+11	source standby flag

Panel C: own changes, bridge, diagnostic and placebos

Schedule	Calls, %	Wait, s	Validity, pp	Excess waiting, patient-h	Relative, %	Added calls without valid interval	More calls / slower / both, h	Shares, calls / slower	Per year, patient-h
July 4 full day (diagnostic)	-5.2 (-7.1 to -3.3)	-64.4 (-84.8 to -41.7)	-0.3	-97.2 (-126.0 to -63.9)	-16.1 (-20.6 to -11.1)	-4%	-31.3 / -70.6 / 4.7		-97
Marathon, route-adjacent own change	+25.9 (17.0 to 37.5)	+194.9 (137.0 to 258.4)	-5.8	+32.5 (21.4 to 48.4)	+81.4 (56.6 to 113.5)	32%	10.6 / 16.8 / 5.1		+33
UN week, UN-adjacent own change	+5.2 (1.1 to 9.6)	+100.3 (68.7 to 132.8)	-2.8	+3.0 (2.0 to 4.0)	+24.6 (16.6 to 33.0)	60%	0.6 / 2.2 / 0.1		+15
UN week, 7,200-s ceiling (bridge)	+6.2 (2.4 to 10.5)	+105.5 (73.2 to 140.6)	-1.5	+3.3 (2.2 to 4.4)	+24.6 (17.8 to 31.4)	28%	0.9 / 2.3 / 0.2		+17
Placebo Sundays 2012, 2020 (near-far)	+5.8	-21.5	-0.4	+0.4	+1.1	8%	2.0 / -1.6 / 0.0		+0
Virtual UN week 2020 (near-far)	+8.4	-33.1	-0.5	-0.1	-1.0	9%	0.6 / -0.6 / -0.0		-0

Panel D: reproduction gates

Gate	Value	Target	Tolerance	Passed
G1 raw N equals published call count in every local hour (170 ZIPs)	1.651428e+07	1.651428e+07	0.0e+00	TRUE
G4 early-month case days	3.960000e+02	3.960000e+02	0.0e+00	TRUE
G4 early-month control days	9.240000e+02	9.240000e+02	0.0e+00	TRUE

Gate	Value	Target	Tolerance	Passed
G4 early-month drug or poisoning IRR (published fepois)	1.086073e+00	1.086073e+00	1.0e-06	TRUE
G3 July 4 fireworks window treated observations	1.440000e+02	1.440000e+02	0.0e+00	TRUE
G3 July 4 fireworks window control observations	5.760000e+02	5.760000e+02	0.0e+00	TRUE
G3 July 4 fireworks window treated rate	1.772708e+02	1.772708e+02	0.0e+00	TRUE
G3 July 4 fireworks window control rate	1.478247e+02	1.478247e+02	0.0e+00	TRUE
G3 Memorial + Labor Day treated observations	5.760000e+02	5.760000e+02	0.0e+00	TRUE
G3 Memorial + Labor Day control observations	2.304000e+03	2.304000e+03	0.0e+00	TRUE
G3 Memorial + Labor Day treated rate	1.530747e+02	1.530747e+02	0.0e+00	TRUE
G3 Memorial + Labor Day control rate	1.624253e+02	1.624253e+02	0.0e+00	TRUE
G2 marathon 2013 active event effect	9.521916e+01	9.521916e+01	1.0e-06	TRUE
G2 marathon 2014 active event effect	4.240497e+01	4.240497e+01	1.0e-06	TRUE
G2 marathon 2015 active event effect	1.842209e+02	1.842209e+02	1.0e-06	TRUE
G2 marathon 2016 active event effect	1.322651e+02	1.322651e+02	1.0e-06	TRUE
G2 marathon 2017 active event effect	1.054147e+02	1.054147e+02	1.0e-06	TRUE
G2 marathon 2018 active event effect	7.876229e+01	7.876229e+01	1.0e-06	TRUE
G2 marathon 2019 active event effect	2.382277e+02	2.382277e+02	1.0e-06	TRUE
G2 marathon 2021 active event effect	1.518908e+02	1.518908e+02	1.0e-06	TRUE
G2 marathon 2022 active event effect	3.196615e+02	3.196615e+02	1.0e-06	TRUE
G2 marathon 2023 active event effect	1.703123e+02	1.703123e+02	1.0e-06	TRUE
G5 UNGA equal-year dT, 7,200-s ceiling, within 15 s of published	1.055474e+02	1.126415e+02	1.5e+01	TRUE

Supplementary Table S47. Calendar-wide scan of other recurring dates. The design and candidate dates were fixed and hashed on 23 October 2026 before any fit to real outcomes, after all published estimates were known. For each of six endpoints (drug or poisoning, assault, mental health, cardiac arrest, burns, all calls) and each candidate, the joint calendar model of Supplementary Table S31 (city-day counts, 2012-2023, January included; year-month strata, weekday-by-month effects, day-of-month terms, New Year's Day, July 4, Memorial and Labor Day lead-lag terms, day-0 terms for five other federal holidays, a temperature cross-basis, precipitation and a physical-hours offset) was refitted with one indicator for the candidate's dates. Candidates were every calendar date outside already-modelled windows (342 dates) and 19 movable or religious observances (Panel B); lunar, Islamic, Hebrew and Hindu dates come from tabular calendars and may differ by a day from local observance. Intervals are clustered by calendar year (t, 11 degrees of freedom), chosen because each candidate occurs once a year and conservative relative to year-month clustering. With 12 clusters, many fits needed covariance corrections (the run log records at least 100), the problem that made year clustering unreliable for the full joint model in Supplementary Table S31, so intervals and q values are approximate screening values. The q values use Benjamini-Hochberg control within each endpoint and family, 2,166 tests in all. A flagged pulse ($q < 0.05$) was refitted on 2024-2025 counts with Central Park weather as a direction check only (two occurrences). The base model reproduced the published day-2 drug or poisoning coefficient exactly, and every indicator was estimable on a simulated outcome. Panel C reports, from the same

model, the five federal holidays that were fitted as adjustment terms and the studied pulses for comparison. Flagged dates are exploratory leads, not confirmed pulses.

Panel A: candidates flagged at q < 0.05

Endpoint	Candidate	Family	Occurrences	Change, % (95% CI)	q	2024-2025, %	Same direction
Assault	31 October (Halloween)	A	12	+46.5 (30.9 to 63.9)	0.0043	+9.7	yes
All calls	Election Day	B	12	-4.8 (-7.4 to -2.0)	0.0340	-6.8	yes
All calls	Easter Sunday	B	12	-3.4 (-5.5 to -1.4)	0.0340	-0.6	yes
All calls	Friday after Thanksgiving	B	12	-4.2 (-6.9 to -1.6)	0.0340	-3.6	yes
Drug or poisoning	Eve of Thanksgiving	B	12	+11.4 (6.7 to 16.3)	0.0036	+20.4	yes
Mental health	Friday after Thanksgiving	B	12	-12.4 (-16.4 to -8.3)	0.0010	-15.5	yes
Mental health	Election Day	B	12	-7.2 (-10.1 to -4.1)	0.0037	-6.8	yes

Panel B: movable and religious observances, smallest q across the six endpoints

Observance	Smallest q (endpoint)	Change, %
Friday after Thanksgiving	0.001 (Mental health)	-12.4
Eve of Thanksgiving	0.0036 (Drug or poisoning)	+11.4
Election Day	0.0037 (Mental health)	-7.2
Easter Sunday	0.034 (All calls)	-3.4
Yom Kippur	0.058 (All calls)	-3.6
Mother's Day	0.1 (All calls)	-3.0
Father's Day	0.12 (Drug or poisoning)	-5.4
Rosh Hashanah	0.13 (All calls)	-2.7
Pride March Sunday	0.15 (All calls)	+2.3
Fall clock change (marathon Sunday)	0.16 (Mental health)	-6.4
Spring clock change	0.17 (Assault)	+26.1
Ramadan (all days)	0.25 (All calls)	-2.1
Super Bowl Sunday	0.33 (Assault)	+12.0
Diwali	0.51 (Drug or poisoning)	+4.6
Eid al-Fitr	0.51 (Drug or poisoning)	+4.2
Puerto Rican Day Parade Sunday	0.64 (Drug or poisoning)	+2.5
Eid al-Adha	0.66 (All calls)	-0.6
Lunar New Year	0.66 (All calls)	-0.8
Passover (day 1)	0.66 (All calls)	-1.1

Panel C: federal holidays fitted as adjustment terms, with the studied pulses for comparison (% change, 95% CI)

Holiday or pulse (day 0)	Assault	Burns	All calls	Cardiac arrest	Drug or poisoning	Mental health
Thanksgiving	+32.9 (5.4 to 67.5)	+70.8 (50.4 to 94.0)	-14.4 (-17.2 to -11.5)	+2.4 (-11.0 to 17.8)	+14.2 (9.6 to 19.0)	-22.8 (-26.9 to -18.6)
Christmas	+44.6 (17.6 to 77.8)	-11.1 (-32.4 to 17.0)	-7.7 (-11.0 to -4.4)	+11.1 (5.4 to 17.1)	+12.8 (5.9 to 20.1)	-18.7 (-21.9 to -15.4)
MLK Day	-8.4 (-30.1 to 20.1)	-30.5 (-49.2 to -5.0)	-9.0 (-10.8 to -7.1)	-5.1 (-14.5 to 5.4)	-3.4 (-9.4 to 2.9)	-12.7 (-15.7 to -9.6)
Presidents Day	+26.5 (0.0 to 59.9)	-3.0 (-26.9 to 28.7)	-10.0 (-11.9 to -8.0)	-11.6 (-15.8 to -7.2)	-1.4 (-6.0 to 3.3)	-13.1 (-16.2 to -10.0)
Columbus Day	+8.0 (-6.9 to 25.3)	-3.0 (-26.2 to 27.5)	-7.3 (-9.7 to -4.8)	-2.5 (-11.2 to 7.0)	-0.2 (-6.3 to 6.2)	-11.3 (-14.3 to -8.1)
studied: day 1	+1.0 (-6.1 to 8.6)	-2.0 (-10.0 to 6.7)	+0.8 (-0.3 to 1.8)	+3.6 (0.6 to 6.8)	+8.2 (6.9 to 9.6)	+0.4 (-1.3 to 2.2)
studied: day 2	-1.1 (-6.5 to 4.7)	+1.9 (-5.8 to 10.2)	+1.2 (0.2 to 2.2)	+3.9 (-0.0 to 7.9)	+8.9 (6.8 to 11.0)	-0.2 (-1.7 to 1.4)
studied: day 3	-1.8 (-7.5 to 4.2)	-2.1 (-10.0 to 6.5)	+1.1 (0.2 to 2.0)	+5.4 (1.6 to 9.3)	+7.6 (5.3 to 9.9)	-0.3 (-2.2 to 1.6)
studied: July 4 (weekday)	+83.4 (57.5 to 113.4)	+89.2 (48.4 to 141.2)	-7.0 (-8.5 to -5.5)	-4.2 (-12.8 to 5.2)	+22.8 (17.3 to 28.4)	-17.8 (-20.6 to -14.8)
studied: Memorial Day	+23.9 (4.9 to 46.2)	+11.6 (-5.3 to 31.4)	-10.2 (-12.5 to -7.8)	-0.5 (-8.9 to 8.8)	+7.6 (1.0 to 14.6)	-19.3 (-22.5 to -16.0)
studied: Labor Day	+66.1 (46.3 to 88.6)	-7.9 (-35.5 to 31.4)	-7.8 (-9.9 to -5.6)	-7.0 (-14.6 to 1.3)	+22.3 (16.1 to 28.9)	-14.0 (-19.1 to -8.7)

Supplementary Table S48. Marathon delay for calls unlikely to come from the event. The design was fixed and hashed on 28 October 2026, after the published marathon estimates were known and before any response time was computed for these subsets. Subsets use the initial call type and severity, assigned when the incident is created. Resident-type calls are call types whose FDNY descriptions make them unlikely to arise from runners or spectators: psychiatric, obstetric and gynaecological, stroke, illness in children under 5, medication reaction, hypertension, dialysis, internal bleeding, violence, fire and environmental, death confirmation and choking. Every ambiguous type (collapse, breathing, cardiac, injury, sick, unknown, altered mental status, seizures, abdominal pain, intoxication, heat and cold, traffic, mass-casualty, standby and special-event codes) is excluded from it. Not-elevated types are call types whose route-adjacent active-window count summed over race days did not exceed the control-Sunday mean, a rule that uses counts only. Life-threatening calls are initial severity segments 1-3. Records flagged as special-event or standby are excluded from every subset except all calls. Estimates are equal-year means of the route-adjacent minus distant change in mean call-to-scene time on the 10 race Sundays relative to their four matched Sundays, with t intervals on 9 degrees of freedom; route-adjacent and distant columns are the mean own changes. All calls reproduce the published active-window estimate exactly. Call types describe the reported condition, not whether the caller lived in the ZIP code. Phases: prior Saturday 21:00-23:59, Sunday 07:00-20:59 (active) and 21:00-23:59 (post).

Subset	Phase	Delay, s (95% CI)	Years positive	Route-adjacent change, s	Distant change, s	Race-day route-adjacent calls
All calls	Prior Saturday	+4 (-79 to 87)	6 of 10	+25	+21	732
All calls	Active	+152 (93 to 210)	10 of 10	+195	+43	3538
All calls	Post	+47 (-34 to 128)	5 of 10	+31	-16	585
Without special-event and standby records	Prior Saturday	+4 (-80 to 87)	6 of 10	+24	+21	732
Without special-event and standby records	Active	+152 (93 to 211)	10 of 10	+195	+43	3538
Without special-event and standby records	Post	+47 (-34 to 127)	5 of 10	+31	-16	585
Resident-type call types	Prior Saturday	+12 (-106 to 129)	3 of 10	+25	+13	93
Resident-type call types	Active	+118 (58 to 179)	8 of 10	+204	+85	410
Resident-type call types	Post	+158 (15 to 301)	8 of 10	+179	+21	82
Call types not elevated on race days	Prior Saturday	-39 (-199 to 121)	4 of 10	-8	+31	77
Call types not elevated on race days	Active	+123 (34 to 213)	9 of 10	+211	+88	322
Call types not elevated on race days	Post	+114 (-29 to 257)	8 of 10	+131	+17	93
Life-threatening (segments 1-3)	Prior Saturday	-28 (-53 to -3)	2 of 10	-22	+6	253
Life-threatening (segments 1-3)	Active	+151 (108 to 194)	10 of 10	+167	+16	1423
Life-threatening (segments 1-3)	Post	+11 (-39 to 61)	4 of 10	+4	-7	194
Event-plausible call types	Prior Saturday	+4 (-86 to 95)	6 of 10	+26	+22	639
Event-plausible call types	Active	+164 (99 to 229)	10 of 10	+199	+35	3128
Event-plausible call types	Post	+36 (-40 to 113)	6 of 10	+16	-21	503

Supplementary Table S49. Three further checks of the street-operation access results. Each design and decision rule was fixed and hashed on 2 November 2026 before estimation, after the published estimates were known. Panel A re-estimates the marathon active-window event effect (route-adjacent minus distant change in call-to-scene time) for the 10 held races with only the standard-time comparison Sundays (2 and 4 weeks after the race), only the daylight-time Sundays (2 and 4 weeks before) or each set of three Sundays, and applies the same design and corrected course classification to the 12 Sundays on which daylight saving time began (2012-2023), with comparison Sundays 2 and 4 weeks either side. Values are equal-year means with t intervals (9

degrees of freedom for races, 11 for spring-forward Sundays). A clock-change explanation would have been supported if the standard-time estimate fell below 50 s or its interval included zero, or if the spring-forward mean reached half the race estimate; neither occurred. Spring-forward Sundays differ from fall-back Sundays in direction and season, and the standard-time Sundays fall later in autumn; the 2012 and 2020 fall-back Sundays without a race are in Supplementary Table S29. Panel B applies the marathon design (corrected course; 34 route-adjacent and 123 distant ZIPs) and the United Nations General Assembly design unchanged to 2024 and 2025, under a design fixed on 2 October 2026 before the 2024-2025 records were extracted. A gate on the candidate-record count failed and was amended on 5 October before any 2024-2025 estimate was computed. Estimates on the earlier course reconstruction (+101 and +172 s) had been seen before the corrected course was fixed, without EMS outcomes, on 19 October; 12 ZIPs newly classified by that course were added from a 2 November extract. Annual marathon effects carry no interval by construction; the pooled incident-level model clusters by ZIP. UN estimates are year-specific with ZIP-and-date-clustered intervals. The prespecified rule required all four primary estimates (marathon active window and UN all hours under 7,200 s, in each year) to be positive; all four were. Panel C refits the Supplementary Table S16 permit-census model without the 13 federal-holiday dates that had a multi-precinct parade or race (Panel D), without Manhattan precincts, without both, and by year; exclusions drop whole precinct-hours, and comparisons stay within borough, date and hour. The prespecified rule classed the estimate without both as not generalizing if its upper bound fell below the published +37.4 s and its point estimate below half of that; it did not generalize. The 1 March 2025 split precedes the FDNY nearest-hospital transport directive of March 2025 by about two weeks. Manhattan-only rows were added after the prespecified results were seen.

Panel A: clock-change checks for the marathon active window

Comparison	Event effect, s (95% CI)	Positive	Distant-ZIP change, s (95% CI)
Race Sundays, all four comparison Sundays (principal)	+151.8 (93.4 to 210.3)	10 of 10	+43.0 (2.8 to 83.3)
Race Sundays, standard-time comparison Sundays only	+156.9 (97.3 to 216.5)	10 of 10	+39.8 (-0.1 to 79.6)
Race Sundays, daylight-time comparison Sundays only	+146.8 (88.4 to 205.1)	10 of 10	+46.3 (4.5 to 88.1)
Race Sundays, each comparison Sunday left out (range of four)	+147.9 to +155.3	10 of 10 in each	+40.5 to +47.8
Spring-forward Sundays, 2012-2023 (placebo)	+1.2 (-21.3 to 23.8)	8 of 12	-16.8 (-42.7 to 9.2)
Spring-forward Sundays, excluding 2020 (placebo)	-2.5 (-25.8 to 20.7)	7 of 11	-6.5 (-20.8 to 7.7)

Panel B: street-closure designs applied to 2024 and 2025, seconds unless stated

Estimand	2024	2025	2012-2023 reference
Marathon, active-window event effect (primary)	+90.6	+164.3	+42.4 to +319.7 (equal-year mean +151.8)
Marathon, pooled incident-level model, 2024 and 2025	+129.8 (60.2 to 199.5)	(pooled with 2024)	n = 23,246 incidents
Marathon, differential valid-interval share (pp)	-4.4	-7.9	
UN week, all hours, intervals under 7,200 s (primary)	+151.4 (38.8 to 264.0)	+65.5 (4.9 to 126.2)	+20.0 to +227.9 (11 in-person years)
UN week, all hours, intervals under 3,600 s	+149.2 (89.0 to 209.4)	+47.4 (1.1 to 93.6)	
UN week, 07:00-20:00, intervals under 7,200 s	+224.8 (68.5 to 381.1)	+33.8 (-20.1 to 87.7)	

Panel C: permit-census estimate for multi-precinct full-street parades and races

Sample	Effect, s (95% CI)	P	Event dates	Exposed precinct-hours
All events (published)	+37.4 (15.8 to 59.1)	0.00115	187	3,706
Without the 13 federal-holiday dates	+30.2 (10.5 to 49.9)	0.0036	174	3,573
Without Manhattan precincts	+12.7 (-12.0 to 37.3)	0.319	134	1,623
Without both (prespecified decision)	+9.2 (-14.8 to 33.3)	0.455	129	1,586
2024	+42.8 (14.7 to 70.9)	0.00382	82	1,690
2025	+30.8 (-4.0 to 65.7)	0.0871	105	2,016
2025, 5 January-28 February	+143.1 (-22.1 to 308.4)	0.0954	9	53
2025, 1 March-31 December	+27.2 (-8.2 to 62.6)	0.136	94	1,954
Without both, 2024	+29.7 (-7.1 to 66.6)	0.12	55	696
Without both, 2025	-9.0 (-45.0 to 27.0)	0.627	74	890
Manhattan only (post hoc)	+61.4 (30.8 to 92.0)	0.000757	135	2,083
Manhattan only, without the federal-holiday dates (post hoc)	+51.4 (24.2 to 78.6)	0.00132	125	1,987

Panel D: the 13 federal-holiday event dates

Date	Events
2024-01-01	Nuestra Senora De La Nube (Parade)
2024-01-15	Martin Luther King Jr. Commemorative Walk 2023 (Parade)
2024-09-02	Labor Day JOuvert Celebration (Parade); West Indian American Day Labor Day Parade (Parade)
2024-10-14	Columbus Day Parade (Parade)
2024-11-11	2024 New York City Veterans Day Parade (Parade); Veterans Day Parade (Parade)
2024-11-28	98th Annual Macys Thanksgiving Day Parade (Parade)
2025-01-01	Nuestra Senora de La Nube (Parade)
2025-01-20	Martin Luther King Jr. Commemorative Walk 2025 (Parade)
2025-05-26	Memorial Day Parade (Parade); STATEN ISLAND MEMORIAL DAY PARADE (Parade)
2025-09-01	Labor Day JOuvert Celebration (Parade); West Indian American Labor Day Parade (Parade)
2025-10-13	Columbus Day Parade (Parade)
2025-11-11	2025 New York City Veterans Day Parade (Parade); Veterans Day Parade (Parade)
2025-11-27	99th Annual Macys Thanksgiving Day Parade (Parade)

Supplementary Table S50. Each schedule's geography. Both designs were fixed and hashed on 11 November 2026 before fitting, after Supplementary Tables S32 and S33 were known. Panel A refits the Supplementary Table S32 stage-1 ZIP models of daily drug or poisoning calls separately in odd and even years (primary) and in 2012-2017 and 2018-2023 (persistence, descriptive). Bivariate random-effects models give the correlation of true ZIP effects between the two halves, for a pulse with itself (own map) and with each other tested pulse (cross-pulse, averaged over both orderings); intervals come from 499 ZIP bootstrap resamples, and several upper limits sit at 1. A pulse had its own geography if the own-map interval and the interval for its excess over the cross-pulse mean both lay above zero; a map whose own-map interval included zero was not stable enough to test. Panel B divides each slope from a multivariate meta-regression of the Supplementary Table S33 stage-1 estimates by the pulse's between-ZIP standard deviation, so pulses with the same geography up to scale have equal values. The five neighbourhood measures are those retained by false-discovery control in Supplementary Table S33, a selection made from that table; log distance to the rebuilt marathon course is the sixth. Panel C tests whether the pulses share one profile (Wald tests using the bootstrap covariance of the scaled differences, with Benjamini-Hochberg control across measures) and whether the calendar maps follow the course; rows labelled post hoc, which exclude New Year's Day, were added after the prespecified results were seen. Panel D gives the share of each pulse's between-ZIP variance explained by the sociodemographic block (income, Black share, public housing), the built-environment block (density, pawnbrokers) and all five, as point estimates. Panel E models marathon-day changes in call-to-scene time at the incident level (active window, 10 races, four comparison Sundays each) with ZIP-by-race and Sunday fixed effects, two-way clustering by race and ZIP and t intervals with 9 degrees of freedom; it is incident-weighted, unlike the equal-year headline. The prespecified rule classed the delay as following the operation rather than residents if the route-adjacent interval lay above zero and the income, Black-share and public-housing intervals lay within 50 s per standard deviation either side of zero; it did. Seventeen of 121 ZIPs in the marathon incident file lack at least one neighbourhood measure and are excluded (786 of 60,653 incidents), among them route-adjacent 11249, which has no polygon in the density source; the calendar models use the 168 ZIPs with all measures. All measures are ecological, attach to incident location and are collinear; licence and facility lists are current snapshots.

Panel A: correlation of true ZIP effects between halves (95% bootstrap interval)

Split	Pulse	Own map	Cross-pulse mean	Own minus cross	Decision
Odd vs even years (primary)	Early month	0.71 (0.23 to 1.00)	0.01 (-0.31 to 0.27)	0.70 (0.10 to 1.27)	own geography
Odd vs even years (primary)	Memorial and Labor Day	0.70 (0.28 to 0.98)	0.21 (0.00 to 0.40)	0.49 (0.08 to 0.80)	own geography
Odd vs even years (primary)	July 4	0.58 (-0.00 to 0.97)	0.07 (-0.16 to 0.29)	0.51 (-0.16 to 0.98)	map not stable enough to test

Split	Pulse	Own map	Cross-pulse mean	Own minus cross	Decision
Odd vs even years (primary)	Late-month dip	0.94 (0.23 to 1.00)			descriptive
2012-2017 vs 2018-2023 (persistence)	Early month	0.99 (0.51 to 1.00)	0.03 (-0.31 to 0.33)	0.96 (0.37 to 1.30)	descriptive
2012-2017 vs 2018-2023 (persistence)	Memorial and Labor Day	0.92 (0.62 to 1.00)	0.20 (-0.05 to 0.43)	0.72 (0.32 to 1.00)	descriptive
2012-2017 vs 2018-2023 (persistence)	July 4	0.39 (-0.17 to 0.80)	0.12 (-0.18 to 0.38)	0.27 (-0.43 to 0.81)	descriptive
2012-2017 vs 2018-2023 (persistence)	Late-month dip	1.00 (0.88 to 1.00)			descriptive

Panel B: scaled slopes, SD of the true ZIP effect per SD of each measure (95% bootstrap interval)

Measure	Early month	Memorial and Labor Day	July 4	New Year's Day
Log median income	-0.11 (-0.32 to 0.13)	-0.10 (-0.36 to 0.11)	0.01 (-0.32 to 0.34)	0.35 (0.12 to 0.55)
Non-Hispanic Black share	0.70 (0.50 to 0.97)	0.46 (0.13 to 0.70)	-0.06 (-0.46 to 0.41)	-0.27 (-0.49 to -0.09)
Public-housing (NYCHA) share	0.19 (0.01 to 0.41)	-0.17 (-0.38 to 0.06)	0.15 (-0.27 to 0.47)	0.09 (-0.10 to 0.28)
Log population density	0.04 (-0.28 to 0.31)	0.12 (-0.13 to 0.40)	-0.56 (-1.09 to -0.09)	0.48 (0.24 to 0.83)
Pawnbrokers per 100,000	-0.05 (-0.16 to 0.44)	-0.04 (-0.85 to 0.04)	0.06 (-0.94 to 0.18)	0.27 (-0.57 to 0.38)
Log distance to the marathon course	-0.14 (-0.45 to 0.14)	0.07 (-0.17 to 0.43)	0.34 (-0.06 to 0.76)	0.06 (-0.16 to 0.31)

Panel C: tests of shared profiles and of the marathon course

Test	Chi-squared	df	P	BH q	Decision
One profile for all four pulses (prespecified)	138.0	15	< 1e-15		distinct neighbourhood profiles
Equal across pulses: log median income	12.3	3	0.00632	0.0105	descriptive
Equal across pulses: Black share	46.1	3	5.43e-10	2.71e-09	descriptive
Equal across pulses: log density	13.9	3	0.00305	0.00762	descriptive
Equal across pulses: public-housing share	6.2	3	0.101	0.126	descriptive
Equal across pulses: pawnbrokers	2.3	3	0.517	0.517	descriptive
Course distance, four pulses jointly (prespecified)	4.2	4	0.385		course association not detected
One profile for early month, Memorial and Labor Day and July 4 (post hoc)	28.6	10	0.00147		post hoc
Early month vs Memorial and Labor Day (post hoc)	11.8	5	0.0375		post hoc

Test	Chi-squared	df	P	BH q	Decision
Early month vs July 4 (post hoc)	19.6	5	0.00148		post hoc
Memorial and Labor Day vs July 4 (post hoc)	13.8	5	0.0173		post hoc

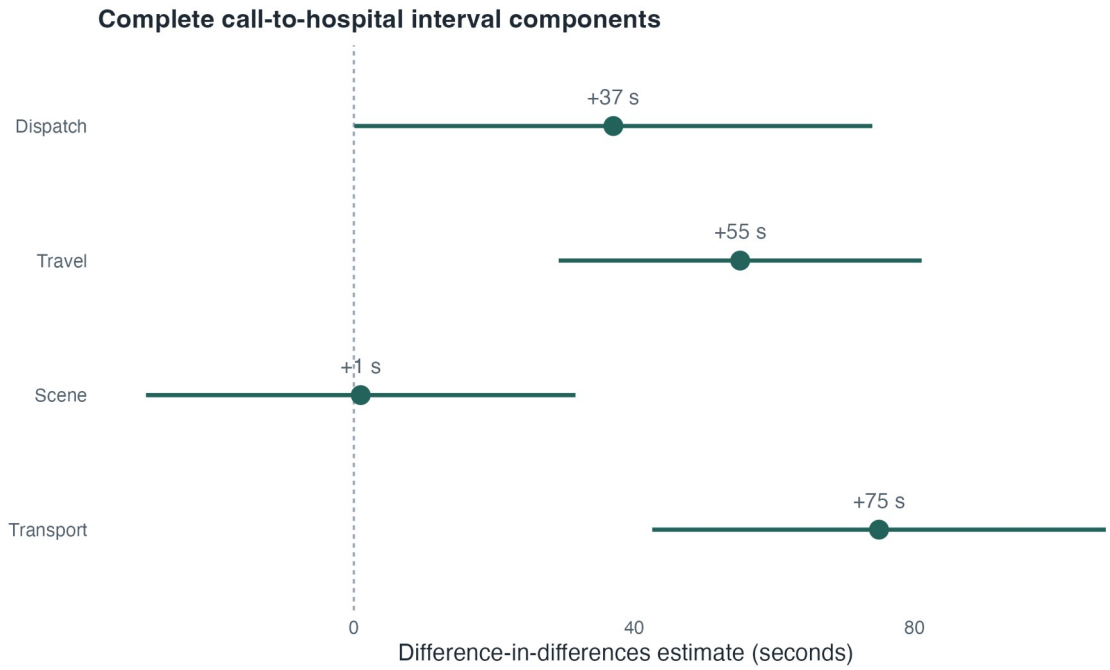
Panel D: share of between-ZIP variance explained

Pulse	Sociodemographic	Built environment	All five
Early month	72%	0%	71%
Memorial and Labor Day	14%	0%	12%
July 4	0%	15%	12%
New Year's Day	29%	20%	48%

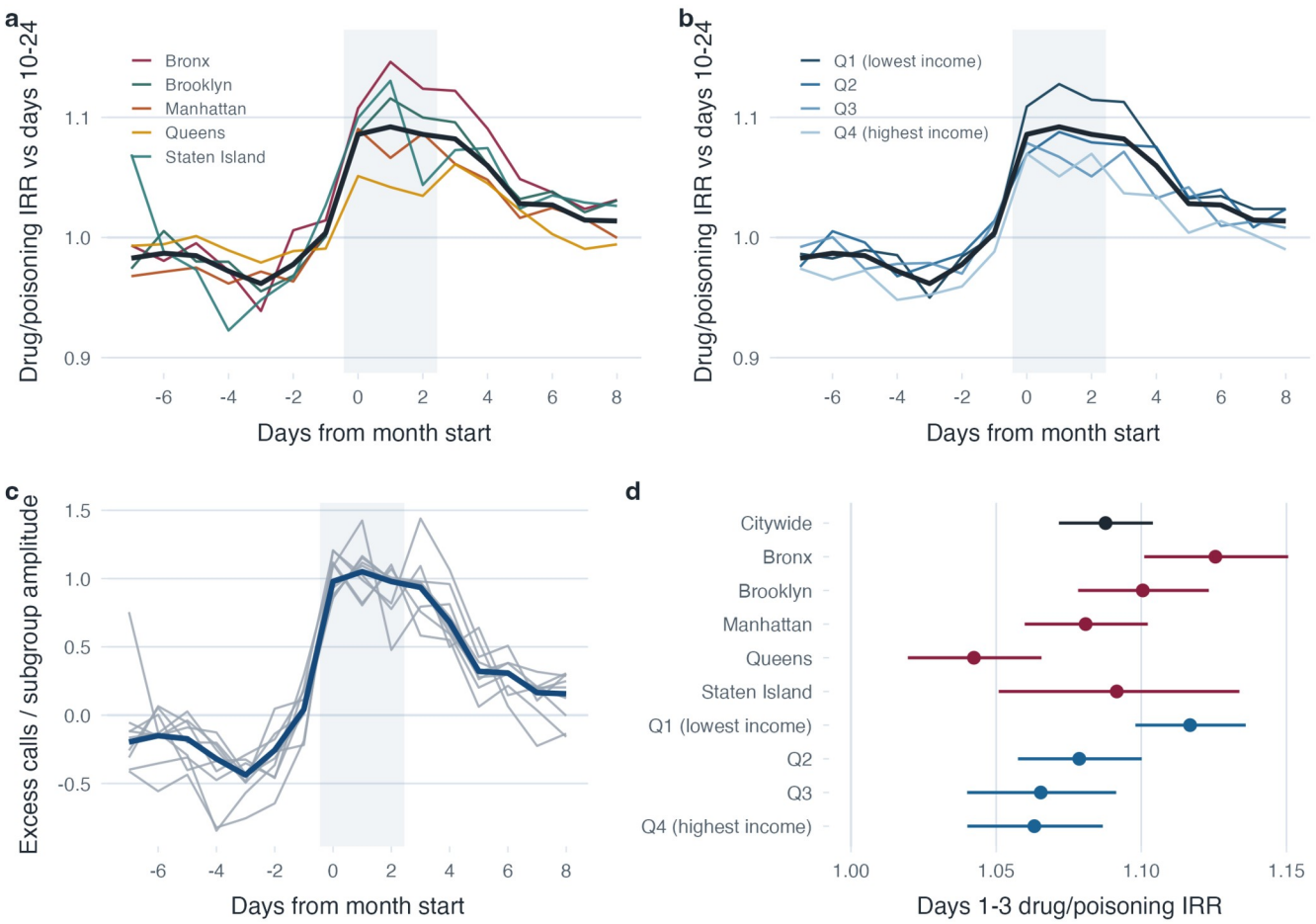
Panel E: marathon delay by route and neighbourhood (59,867 incidents in 104 ZIPs; decision: follows the operation, not the residents)

Term	Marathon-day change, s (95% CI)	P	BH q
Route-adjacent ZIP (within 1 km, vs beyond 5 km)	121.6 (33.1 to 210.0)	0.0125	
Log median income, per SD	1.7 (-35.7 to 39.1)	0.922	0.981
Non-Hispanic Black share, per SD	13.2 (-1.3 to 27.6)	0.0694	0.206
Public-housing share, per SD	14.5 (-2.7 to 31.7)	0.0885	0.206
Log population density, per SD	32.2 (-3.0 to 67.4)	0.0681	0.206
Pawnbrokers, per SD	1.1 (-28.4 to 30.6)	0.936	0.981
Distance to the nearest emergency department, per SD	-1.4 (-16.7 to 14.0)	0.845	0.981
Line to the nearest emergency department crosses the course	1.2 (-106.6 to 108.9)	0.981	0.981

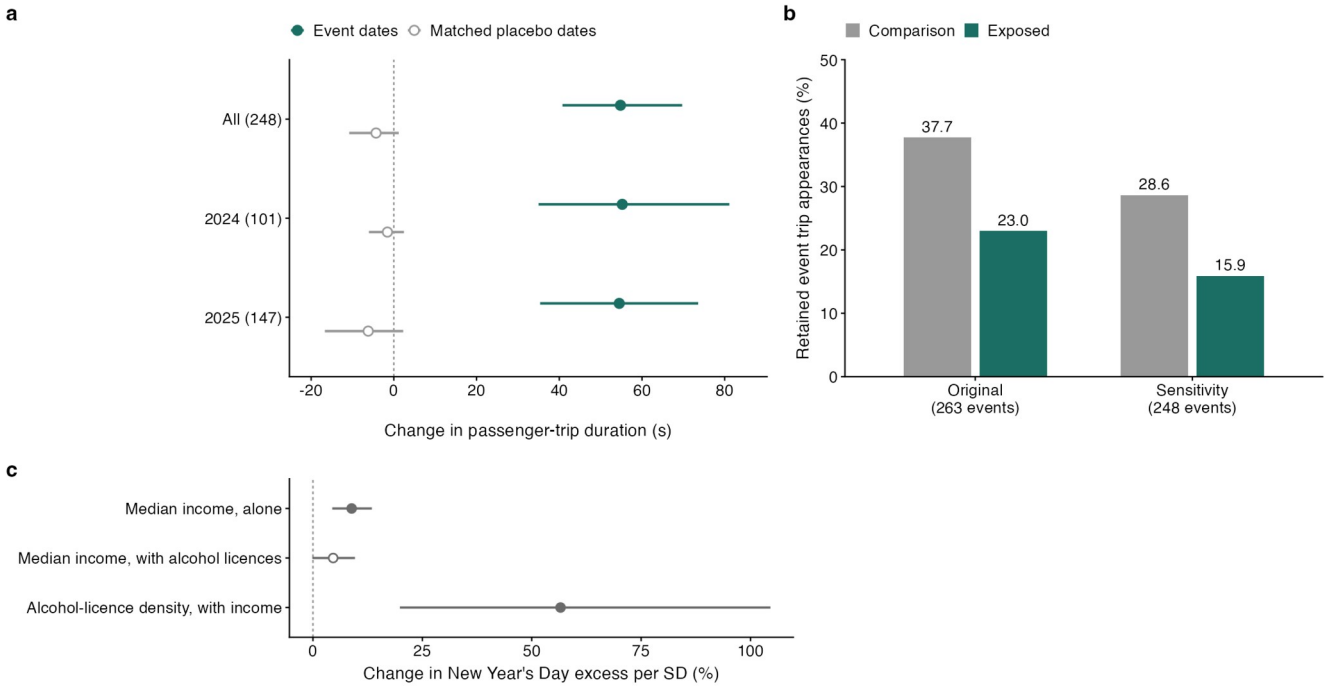
Supplementary Figure S1. Marathon component-specific interval estimates. Separate dispatch, travel, scene and transport models start from the secondary completed-transport sample and apply each component's positive, less-than-one-hour validity rule. Points and normal-reference 95% confidence intervals describe different component-specific samples; the estimates are not additive and do not decompose the complete call-to-hospital interval (Extended Data Table 2).



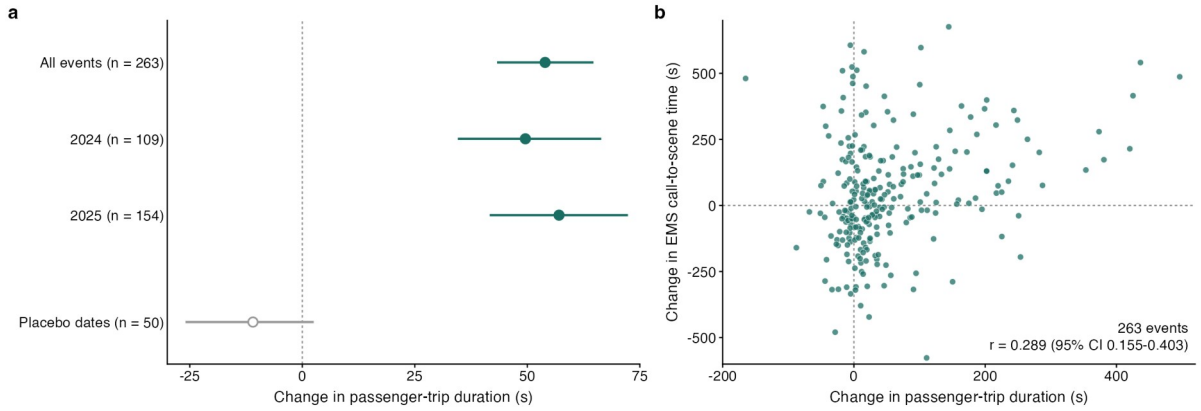
Supplementary Figure S2. The early-month drug or poisoning pulse has a conserved shape across boroughs and income strata. Event-time drug or poisoning incidence rate ratios around the month boundary (day 0 = first of the month; baseline days 10-24; January anchors excluded), by borough (a) and ZIP income quartile (b), with the citywide curve in bold and the days 1-3 exposure window shaded. (c) The same curves rescaled by each subgroup's days 1-3 amplitude align with the citywide shape (mean correlation 0.96, range 0.83-0.99; one-sided circular-shift permutation P = 0.00001): an abrupt rise on day 1, a peak on day 2 and decay to baseline within roughly a week. (d) Days 1-3 amplitudes with 95% confidence intervals (Supplementary Table S7).



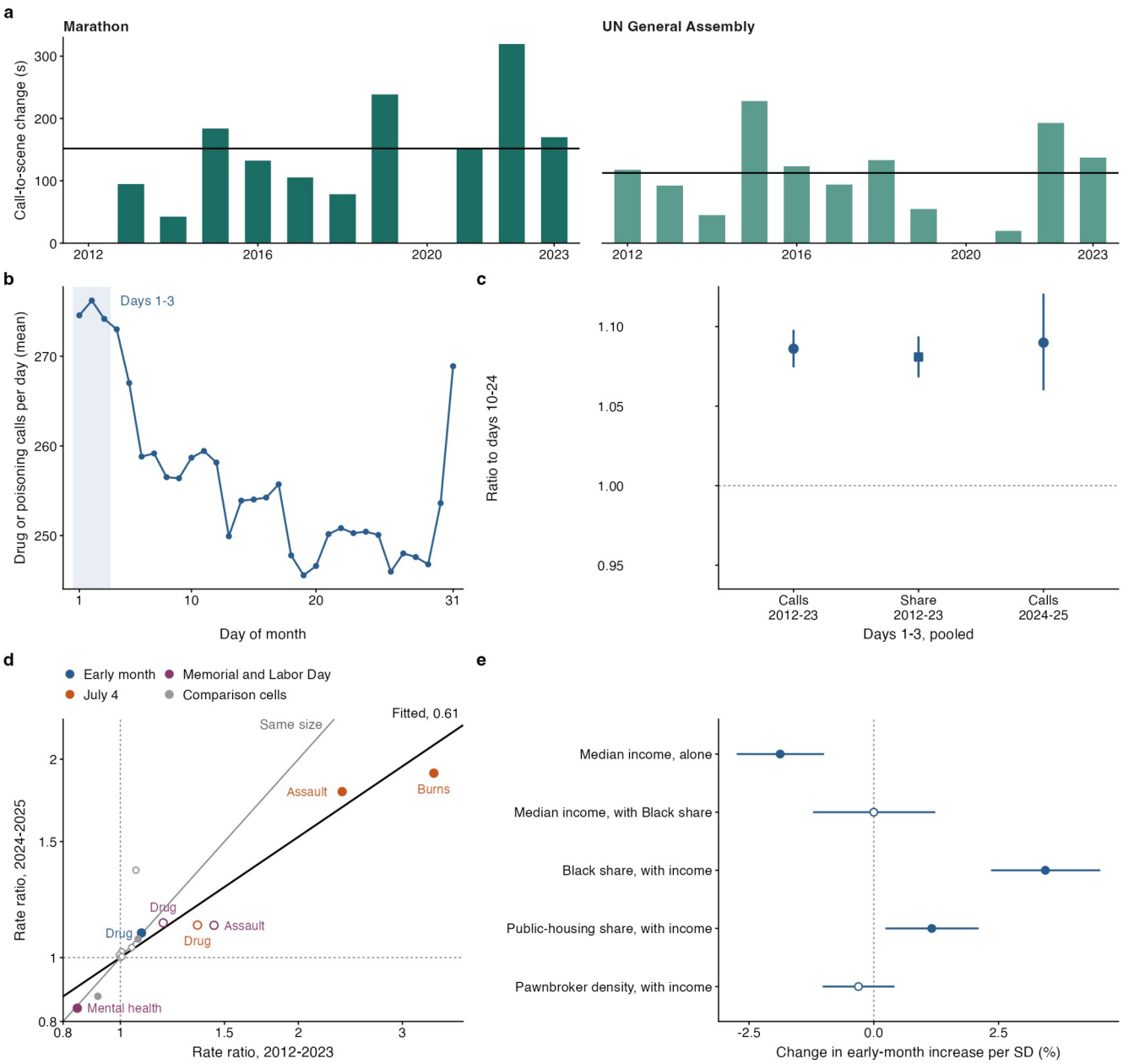
Supplementary Figure S3. Passenger-trip sensitivity, trip retention and New Year's Day moderators. Panels a and b accompany the passenger-trip results in Supplementary Fig. S4, and panel c adds the New Year's Day rows to the moderators in Supplementary Fig. S5e. (a) The passenger-trip sensitivity sets candidate cells and weights using three comparison dates and reserves the fourth as the placebo date. Event and placebo dates use identical observed cells, with at least one valid trip per cell on each date. Supported folds are averaged within events, then events equally. Intervals use 1,999 event-date bootstrap draws; event and placebo share 248 events on 157 dates, including 101 events in 2024 and 147 in 2025. (b) Retained valid event-trip appearances divided by eligible valid appearances, separately by exposure arm. The original and sensitivity cohorts differ. Trips can appear in multiple event comparisons and, in the sensitivity, multiple supported folds; these are not unique citywide trip counts. Durations remain unobserved for missing trips. Supplementary Tables S26 and S28 report complete support ledgers. (c) Change in the New Year's Day excess per standard deviation of median income and alcohol-licence density, alone and with joint adjustment, from the two-stage multivariate meta-analysis of 170 ZIPs (Supplementary Table S33). Filled symbols mark 95% intervals that exclude 0 and open symbols intervals that include 0.



Supplementary Figure S4. Passenger-transport delays during permitted street events. (a) Original equal-event trip-duration contrasts and 95% date-cluster bootstrap intervals; n gives event counts. The placebo covers 50 of 263 events under its stricter one-day support rule; the event estimate in those same 50 is +70.3 s (40.6-100.7; Table 1). Filled points mark intervals that exclude 0; the open placebo interval includes it. (b) All 263 independently estimated transport/EMS event contrasts. EMS call-to-scene time starts at incident creation and includes dispatch. The Pearson correlation and interval describe association, not mediation. Supplementary Fig. S3a,b shows the sensitivity using comparison-date weights with matched placebos and the retained trip shares; Supplementary Tables S26 and S28 report complete support ledgers.



Supplementary Figure S5. Panels supporting the main figures. (a) Annual point estimates for the active-window marathon and all-hours in-person UN General Assembly contrasts; horizontal lines mark the equal-year summaries in Fig. 1c. The marathon was not held in 2012 or 2020, and the 2020 General Assembly met virtually. (b) Descriptive mean daily drug or poisoning calls by day of month across February-December 2012-2023; days 29-31 occur in fewer months, so their raw means are not month-standardized, and day 31 includes 31 December (New Year's Eve). (c) On the scale of Fig. 2b, the primary 2012-2023 call-count model (396 matched strata), category-share model (396 strata) and prespecified 2024-2025 call-count model (66 matched strata). Count models report incidence rate ratios (circles); the share model reports the adjusted odds ratio for drug or poisoning versus all other calls (square), comparing days 1-3 with matched controls, which is not a percentage-point change in share. (d) Rate ratios in 2012-2023 against 2024-2025 for all 15 schedule-by-endpoint cells, on log scales; the grey line marks equal size and the black line the fitted calibration slope across the seven prespecified cells (0.61; Supplementary Table S39). Coloured, labelled points are the seven prespecified cells and grey points the eight comparison cells; filled points mark 2024-2025 95% confidence intervals that exclude 1. The Spearman correlation was 0.914. (e) Change in the early-month pulse size per standard deviation of each neighbourhood characteristic from a two-stage multivariate meta-analysis of 170 ZIPs; "with" denotes joint adjustment, and public-housing share is the New York City Housing Authority share. Neighbourhood measures are ecological and attach to incident location (Supplementary Table S33). In b-e, filled symbols mark 95% intervals that exclude the null and open symbols intervals that include it.


a
Marathon
UN General Assembly
Call-to-scene change (s)
300
200
100
0
2012
2016
2020
2023
b
Drug or poisoning calls per day (mean)
Days 1-3
270
260
250
1
10
20
31
Day of month
c
Ratio to days 10-24
1.10
1.05
1.00
0.95
Calls 2012-23
Share 2012-23
Calls 2024-25
Days 1-3, pooled
d
Early month
Memorial and Labor Day
July 4
Comparison cells
Same size
Fitted, 0.61
Assault
Burns
Drug
Drug
Assault
Drug
Mental health
Rate ratio, 2024-2025
2
1.5
1
0.8
0.8
1
1.5
2
3
Rate ratio, 2012-2023
e
Median income, alone
Median income, with Black share
Black share, with income
Public-housing share, with income
Pawnbroker density, with income
-2.5
0.0
2.5
Change in early-month increase per SD (%)