Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Emergency medical services as sentinel infrastructure for climate-health surveillance

Chris C. Lim
chrislim@arizona.edu

University of Arizona
Jonathan Race
University of Utah
Honghyok Kim
University of Illinois Chicago
Ran Wei
University of California Riverside
Xiaojiang Li
University of Pennsylvania
Lynn Gerald
University of Illinois Chicago
Chen Chen
University of Utah
N Mann
University of Utah

Article

Keywords: emergency medical services, climate change, heat, distributed lag non-linear model, health equity, prehospital care, prehospital monitoring

Posted Date: March 10th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-9003120/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: There is **NO** Competing Interest.

Emergency Medical Services as Sentinel Infrastructure for Climate-Health Impacts

Chris C. Lim[1]*, Jonathan A. Race[2], Honghyok Kim[3], Ran Wei[4], Xiaojiang Li[5], Lynn Gerald[6], Chen Chen[7], N. Clay Mann[2]

[1]Mel & Enid Zuckerman College of Public Health, The University of Arizona, Tucson, AZ, USA

[2]Department of Pediatrics, School of Medicine, University of Utah, Salt Lake City, UT, USA

[3]School of Public Health, University of Illinois Chicago, Chicago, IL, USA

[4]School of Public Policy, University of California Riverside, Riverside, CA, USA

[5]Department of City and Regional Planning, University of Pennsylvania, Philadelphia, PA, USA

[6]Office of Population Health Sciences, University of Illinois Chicago, Chicago, IL, USA

[7]Department of Population Health Sciences, University of Utah, Salt Lake City, UT, USA

*Corresponding author

Contact:

Chris C. Lim

1295 N Martin Ave, Tucson, Arizona, 85724

Mel & Enid Zuckerman College of Public Health, The University of Arizona

chrislim@arizona.edu

512-913-6114

Abstract Word Count: 150 | **Manuscript Word Count:** ~3,900

ABSTRACT

Emergency medical services (EMS) operate at the front line of climate-health impacts, yet how weather changes demand and response capacity is unclear. We analyzed 16.5 million EMS incidents in New York City (2012–2023) across 19 endpoints and 6 response-time components, identifying three dimensions of weather-associated stress: demand surges, operational degradation, and information degradation. Heat increased call volume by 10% and cumulative response time by 65 seconds (+21 per call); dispatch time showed the largest increase (+42.6%). Cold reduced demand and response time, though cardiac arrest increased at winter extremes. Precipitation reduced volume but increased respiratory and obstetric emergencies, with travel-dominated delays. Heat increased non-specific classifications and severity under-triage for symptom-overlapping conditions but not cardiac arrest, consistent with information constraints at the caller–dispatcher interface. Heat-related delays were 2.4× larger in low- versus high-income neighborhoods. These results position EMS as sentinel infrastructure and identify dispatch protocols and travel infrastructure as adaptation targets.

Emergency medical services are the first point of clinical contact for acute health events, yet they remain largely underexamined in climate–health research. For time-sensitive conditions such as cardiac arrest, survival depends critically on response speed. Unlike hospitals, EMS operates directly within the environmental conditions generating patient demand, exposed to the same heat, precipitation, and compound hazards that drive the emergencies they are dispatched to treat. This creates potential for a dual burden not captured by hospital-based surveillance: extreme weather may simultaneously increase demand for prehospital care and degrade the system's capacity to deliver it. As climate change intensifies extreme heat events and compound weather hazards globally (Perkins-Kirkpatrick & Lewis, 2020; Zscheischler et al., 2018), understanding whether and how these pathways operate is essential for protecting populations whose survival depends on minutes.

EMS occupies a distinct position at the interface between environmental exposures and hospital-based care, with three properties relevant to climate–health monitoring. First, temporal resolution: EMS captures system stress as events unfold before patients reach emergency departments, providing an early window for public health response and resource mobilization. Second, operational granularity: EMS data record discrete response intervals (dispatch [call-processing], activation, travel, scene, transport), enabling localization of where delays arise, information that is obscured in aggregate hospital metrics. Third, direct survival relevance: for time-sensitive conditions such as cardiac arrest, survival decreases by approximately 7–10% for each minute of delay to defibrillation (Morgan et al., 2020; Riva & Hollenberg, 2021; Ruiz de Gauna et al., 2014). Although recent studies have linked temperature to EMS call volume (Seong et al., 2023), they have not distinguished whether extreme weather overwhelms EMS through demand-mediated congestion or is associated with degraded performance through independent operational pathways, a distinction with direct implications for adaptation strategy. New York City (NYC), which operates one of the largest municipal EMS systems in the United States (~1.4 million annual 911-dispatched incidents during the study period) and maintains detailed prehospital data infrastructure, provides a well-suited setting to evaluate these questions at scale.

Here we analyze 16.5 million EMS incidents in New York City from 2012 to 2023 using distributed lag non-linear models across 19 clinical endpoints and 6 response-time components. We combine response-time decomposition to localize where delays manifest, queueing diagnostics to distinguish demand-driven congestion from direct capacity degradation, and classification triangulation to identify information-constraint pathways. Heat simultaneously amplified demand (+10% call volume) and eroded response capacity (+11.7% cumulative response time, +65 seconds at 30°C versus 20°C), with delay concentrating in dispatch while precipitation delays shifted to travel — patterns consistent with distinct operational bottlenecks not fully explained by congestion. Severity under-triage increased for symptom-overlapping conditions but remained null for cardiac arrest, consistent with information constraints at the caller–dispatcher interface. Response-time effects were 2.4-fold larger in low-income neighborhoods. The results suggest three dimensions of weather–EMS association (demand surges, operational degradation, and information degradation), each with different adaptation implications.

RESULTS

Study Population and Exposure Characteristics: We analyzed 16.5 million EMS incidents in New York City from January 1, 2012 through December 31, 2023, comprising 17.9 million ZIP-hour observations (170 ZIP codes × 105,192 hours; 169 ZIP codes with valid income data for stratified analyses). The study population included all emergency calls dispatched

through the NYC 911 system. Mean hourly temperature was 12.8°C (range -18.0 to 37.9°C); 2.2% of hours exceeded 30°C (approximately 2,300 hours), 9.5% were below 0°C, and 18% of ZIP-hours had measurable precipitation. Mean daily call volume was 3,768 incidents (SD 360), with cardiac arrest comprising 1.9% of calls (approximately 72 per day), respiratory emergencies 9.8%, mental health 10.2%, cardiovascular 8.0%, unconscious/syncope 5.6%, and non-specific illness 22.6%. The study population covered a wide socioeconomic range: ZIP code median household incomes ranged from $21,447 to $224,063, with 57.3% non-white population citywide (population-weighted). Income-stratified analyses used 169 ZIP codes with valid American Community Survey (ACS) income data (Table 1). Sample sizes differ across analyses because of nested analytic structure: 105,192 city-hours, 17.9 million ZIP-hours, and 16.5 million incidents represent the same underlying data at different levels of aggregation, with smaller analytic samples reflecting response-time timestamp availability (9.1 million ZIP-hours) or income data coverage (169 of 170 ZIP codes). Figures present 10 core endpoints selected for clinical time-sensitivity, mechanistic contrast, or equity relevance (total calls, cardiac arrest, stroke, unconscious/syncope, respiratory, mental health, obstetric, pediatric, non-specific illness, assault); full 19-endpoint results appear in Extended Data. Six potential endpoints were excluded a priori (see Methods).

Results report two complementary response-time estimands. *System-level burden* (city-hour DLNM, cumulative 24-hour lag) captures both direct and lagged pathways, including prior-hour congestion and crew fatigue. *Per-call effects* (incident-level, concurrent exposure) isolate direct weather associations on individual calls. The ~3-fold difference between cumulative (+65 seconds) and per-call (+21 seconds) heat effects on response time is consistent with lagged system-level pathways contributing the majority of delay, though differences in model specification also contribute. Call-volume analyses use only the cumulative estimand; response time is reported under both frameworks. Precipitation response-time estimates use ZIP-hour DLNM rather than city-hour DLNM because precipitation varies more spatially than temperature across the city (Supplementary Table S17); magnitudes are not directly comparable across analytic units.

Temperature Effects on EMS Demand and Response Capacity: Heat (30°C vs 20°C) was associated with increased overall demand (rate ratio [RR] 1.10, 95% confidence interval [CI] 1.10-1.11), translating to approximately 394 additional calls per day citywide (from unrounded cumulative RR 1.1047 × baseline daily mean of 3,768). Among core endpoints, unconscious/syncope showed the largest effect (RR 1.32, 95% CI 1.29-1.35), followed by pediatric emergencies (RR 1.24, 95% CI 1.17-1.31), cardiac arrest (RR 1.22, 95% CI 1.18-1.26), respiratory emergencies (RR 1.16, 95% CI 1.14-1.18), and stroke (RR 1.13, 95% CI 1.08-1.18) (Figure 2A). Non-specific illness (22.6% of calls) increased 14% at 30°C. Effects intensified non-linearly at extreme heat (35°C) (Figure 3; Supplementary Table S6). Cold (0°C vs 20°C) showed the opposite demand pattern: total call volume decreased 13% (RR 0.87, 95% CI 0.87-0.88), with assault (-34%) and unconscious/syncope (-27%) showing the largest reductions. However, cardiac arrest diverged: while moderate cold modestly reduced cardiac arrest calls (RR 0.94 at 0°C), extreme cold increased them (RR 1.06, 95% CI 1.01-1.10 at -10°C), consistent with distinct cardiovascular risk at winter extremes.

Heat was also associated with reduced response capacity. In the city-hour DLNM (cumulative 24-hour lag), total response time increased by 65 seconds (+11.7%, model-based 95% CI +10.4 to +13.0%, widening to approximately +8% to +15% after adjusting for residual autocorrelation [ACF(1) = 0.77; Methods]; baseline at 20°C: 559 seconds) at 30°C versus 20°C. Baselines are model-predicted values at the 20°C reference and differ from unconditional means in Table 1, which average across all temperatures. Comparable autocorrelation affects all response-time models (Supplementary Table S14). In

incident-level models, the per-call heat association was +21 seconds (+3.7%) (Figure 2B), consistent with lagged system-level pathways contributing the majority of cumulative delay. Among components, dispatch showed the largest relative increase (+42.6%, +28 seconds from baseline 66 seconds; autocorrelation-adjusted CI approximately +27% to +58%) while travel showed an identical absolute increase (+28 seconds, +6.1% from baseline 459 seconds); activation was smaller (+4.1%). Components were modeled separately, so their changes do not sum exactly to total. Post-arrival components showed negligible heat effects (Figure 1B; Figure 4). Cold showed the opposite operational pattern: total response time decreased 14.4% (-80.7 seconds) at 0°C, with dispatch showing the largest reduction (-37.7%, -24.9 seconds) and travel also faster (-9.2%, -42.1 seconds), though scene time increased +3.1% (+39.3 seconds), consistent with providers spending longer on-scene in cold conditions (Figure 4A).

Precipitation Effects on Call Volume and Response Time: Unlike heat, precipitation was not associated with uniform demand amplification. At heavy rain (5 mm vs 0 mm), total call volume declined by 9.9% (RR 0.90, 95% CI 0.88-0.92), but endpoint-specific effects diverged. Respiratory calls increased 9.2% (RR 1.09, 95% CI 1.05-1.14), cardiovascular calls increased 5.6% (RR 1.06, 95% CI 1.01-1.10), obstetric calls increased 29.8% (RR 1.30, 95% CI 1.19-1.42), and pediatric calls increased 25.7% (RR 1.26, 95% CI 1.10-1.43).

In contrast, several categories decreased sharply: assault (RR 0.37, 95% CI 0.32-0.44), poisoning/overdose (RR 0.63, 95% CI 0.60-0.67), unconscious/syncope (RR 0.51, 95% CI 0.48-0.54), trauma (RR 0.91, 95% CI 0.88-0.95), and non-specific illness (RR 0.90, 95% CI 0.88-0.93). Effects were larger at higher precipitation levels (Figure 3).

Precipitation was also associated with substantial operational delays, with contrasting profiles to heat. Heavy precipitation (5 mm vs 0 mm) increased total response time by approximately 24–27% in non-winter seasons and by 88% during winter (December–February) (Supplementary Table S16). Whereas heat delays were dispatch-dominated (+42.6%, city-hour DLNM), precipitation delays (ZIP-hour DLNM) within response time showed the largest absolute increase in travel time (+23.1%, approximately +104 seconds given travel's longer baseline) despite dispatch showing the largest relative increase (+63.1%, +41 seconds; Figure 1B). Heat and precipitation estimates derive from different analytic units (city-hour and ZIP-hour DLNM, respectively; Supplementary Table S17), but the contrast in which component dominates is consistent across model specifications. Post-arrival stages (not included in the response time metric) showed large absolute increases: scene time (+22.2%, +285 seconds) and transport (+7.2%, +51 seconds). The winter amplification occurred independently of call volume changes.

Demand elasticity of response time remained within a narrow range (8.8–10.1%) across weather conditions, and heat effects persisted during low-utilization hours (Extended Data Figure 3). Under compound hot-wet conditions (30°C, 5 mm), response time increased 25.2%, approximately matching the product of individual heat (+11.7%) and precipitation (+12.1%, annual city-hour DLNM; seasonal ZIP-hour estimates are larger at +24–27% non-winter, reflecting finer spatial resolution and seasonal heterogeneity) effects (1.117 × 1.121 = 1.252). The temperature×precipitation interaction surface was significant (LRT $P < 0.001$; Extended Data Table 1); however, this test detects departures from independence across the full exposure surface, not specifically at the hot-wet combination. Departures from multiplicative scaling were concentrated in cold-wet conditions; at the hot-wet compound most relevant to climate adaptation, heat and precipitation effects combined approximately multiplicatively.

Classification and Severity Patterns During Heat: Heat was associated with degraded information quality at the caller–dispatcher interface. Non-specific classification (calls initially coded as "sick," "unknown," or "general illness"; distinct from the non-specific illness endpoint) increased at 30°C for cardiac arrest (RR 1.24, 95% CI 1.05-1.46) and cardiovascular (RR 1.12, 95% CI 1.05-1.19). Calls coded as "caller has no patient medical information" increased during heat across all endpoints examined (e.g., cardiac arrest: 3.10% to 3.70%). Severity under-triage (initial priority lower than final) increased for obstetric emergencies (RR 1.31, 95% CI 1.13-1.51) and stroke (RR 1.22, 95% CI 1.06-1.41), while cardiac arrest, pre-specified as a negative control, showed null under-triage (RR 0.99, 95% CI 0.95-1.04) (Figure 5; Extended Data Table 2; Extended Data Figure 4). The increased non-specific *initial classification* for cardiac arrest (RR 1.24) but null *severity assessment* (RR 0.99) is consistent with structured dispatch protocols overcoming caller ambiguity for unambiguous presentations. Effects intensified at 35°C: stroke under-triage reached RR 1.46 (95% CI 1.11-1.92) while cardiac arrest remained null (RR 0.98). Severity code completeness was temperature-independent ($P = 0.42$). Provider-level correct identification remained near unity across endpoints (Extended Data Table 2), localizing degradation to the caller-dispatcher interface. Unconscious/syncope served as a second negative control: despite +32% demand surge, severity under-triage was null (RR 0.97, 95% CI 0.92-1.03). Cold temperatures did not produce analogous degradation: unadjusted under-triage rates during cold (<0°C) were similar to or lower than the reference for cardiovascular (-0.6 percentage points [pp]), respiratory (-0.5 pp), and stroke (-0.4 pp). This asymmetry is consistent with cold symptoms being clinically distinct, whereas heat symptoms overlap with cardiac, stroke, and syncope presentations. Downstream of these classification patterns, heat was associated with lower hospital transport probability (odds ratio [OR], where OR < 1 indicates reduced transport; OR 0.88 at 30°C, OR 0.80 at 35°C), with reductions for ambiguous presentations (unconscious/syncope OR 0.84, 95% CI 0.80-0.87; −4.2 percentage points from a baseline transport rate of 62%), respiratory (OR 0.80, 95% CI 0.77-0.84), and cardiovascular (OR 0.87, 95% CI 0.83-0.92) endpoints. Cardiac arrest also showed reduced transport (OR 0.79, 95% CI 0.73-0.85), but from a baseline transport rate of 50%, substantially lower than other endpoints (62–86%), reflecting that many cardiac arrest patients are pronounced dead at scene. Stroke transport was not statistically significantly reduced (OR 0.91, 95% CI 0.79-1.05). At 35°C, overall transport probability decreased further (OR 0.80), with unconscious/syncope reaching OR 0.77 (95% CI 0.72-0.83).

Disproportionate Operational Burden by Income: Income-stratified DLNM analysis showed similar relative heat effects on call volume across income quartiles (RR 1.10–1.11; heterogeneity $P = 0.44$; RRR Q1/Q4=1.00), but the pre-specified Q1-Q4 response-time contrast showed a 2.4-fold difference in heat effects (Q1: +20.35% vs Q4: +8.55%, from separate within-quartile DLNMs; no formal between-quartile CI is available because quartile models were fitted independently). This contrast should be interpreted cautiously: the response-time pattern was not monotonic across intermediate quartiles (Q3: +14.30% > Q2: +10.22%), indicating that this is a Q1-versus-Q4 disparity rather than a dose-response income gradient. Because Q1 ZIP codes cluster geographically (predominantly Bronx and central Brooklyn), the 2.4-fold ratio may partly reflect borough-level differences in ambulance depot density, hospital proximity, or traffic patterns rather than income per se. Baseline call rates were higher in Q1 than Q4 (57.9 vs 46.8 per 100,000/day), and baseline response times were approximately 13% longer (615 vs 546 seconds). The composite Burden Index (normalized to Q4 = 1.00) was highest in Q1 (1.39×), with Q2 at 0.87× and Q3 at 0.60× (Extended Data Table 3); Q3 falling below the Q4 reference further illustrates that the income–burden relationship is not monotonic and likely reflects borough-level operational geography rather than a clean income gradient. Precipitation showed distinct income gradients for climate-sensitive medical

conditions: heavy rain (5 mm) increased respiratory calls in lower-income neighborhoods (Q1: RR 1.10; Q2: RR 1.13) but not in higher-income neighborhoods (Q3–Q4 not significant), and pediatric calls increased significantly only in the lowest-income quartile (Q1: RR 1.24, 95% CI 1.07-1.44). Cardiovascular and obstetric emergencies showed similar income-graded precipitation effects (Extended Data Figure 2).

Seasonal and Time-of-Day Effect Modification: Heat effects on total calls were largest outside peak summer (winter RR 1.17; spring RR 1.15; summer RR 1.13; fall RR 1.11; Supplementary Table S15), with heat effects declining monotonically from winter to fall. This seasonal gradient was also apparent for individual endpoints in within-season percentile contrasts (95th vs 50th percentile; Supplementary Tables S7-S8), which yield more conservative estimates by comparing narrower temperature ranges within each season. Using the fixed 30°C contrast, the gradient was amplified for cardiac arrest (spring RR 1.86, 95% CI 1.62-2.14; summer RR 1.06, 95% CI 1.00-1.13), though the spring estimate draws from limited hours exceeding 30°C (primarily late May) and should be interpreted with caution. Respiratory effects remained elevated across seasons. Cold temperatures showed a contrasting pattern: moderate cold (-5°C vs 20°C) was associated with reduced total call volume (RR 0.87, 95% CI 0.86-0.88), but cardiac arrest increased during extreme winter cold (-7°C vs 2°C; RR 1.06, 95% CI 1.02-1.10; Supplementary Tables S6, S8). Precipitation–response time effects also varied seasonally: winter precipitation produced the largest delays (+88% at 5 mm vs 0 mm, compared with approximately +24–27% in non-winter seasons; Supplementary Table S16), consistent with snow and ice compounding travel-time effects.

Time-of-day stratification showed that apparent delayed heat effects in pooled models were consistent with temporal mixing: heat effects were strongest during morning hours (RR 1.16) and cardiac arrest effects were most pronounced at night (RR 1.20). By contrast, precipitation retained delayed 12–24-hour lag structures across all time strata.

DISCUSSION

These findings suggest three dimensions of weather-associated EMS stress: demand surges, operational degradation, and information degradation, with disproportionate burden in lower-income neighborhoods. This framework extends prior evidence linking heat to increased ambulance dispatches, where meta-analytic estimates are typically 9–11% (Turner et al., 2013; Alessandrini et al., 2011; Kotani et al., 2018; Yuan et al., 2023; Cheng et al., 2016; Ma et al., 2015; Wang & Lin, 2020; Xu et al., 2023). Earlier work emphasized volume rather than operational intervals (Cui et al., 2020; Zhan et al., 2020; Seong et al., 2023). Here, dispatch time was the most heat-sensitive component; classification, severity, and transport patterns were consistent with a symptom-overlap mechanism; and queueing diagnostics were compatible with partially independent demand and capacity pathways. Heat effects for cardiac arrest (+22%) and pediatric calls (+24%) were larger than prior studies, potentially reflecting NYC heat-island intensity and 12 years of hourly data.

The dual burden pattern (~10% higher demand and 11.7% slower cumulative response at 30°C) extends prior heat-dispatch evidence into the operational domain. The cumulative DLNM estimate (+65 seconds) captures 24-hour system burden, whereas the concurrent per-call estimate (+21 seconds) reflects direct weather effects; their difference implies lagged contributors such as prior-hour congestion and crew fatigue. For cardiac arrest (Morgan et al., 2020; Riva & Hollenberg, 2021), even +21 seconds may be clinically meaningful at NYC scale. Heat delays were dispatch-dominated (+42.6%), a modifiable information-processing stage; precipitation delays were travel-dominated, implying different operational responses. Cold temperatures produced the opposite pattern: response time decreased 14.4%, with dispatch

showing the largest reduction (-37.7%), while scene time increased (+3.1%). The bidirectional dispatch effect (heat +42.6%, cold -37.7%) suggests call-processing speed is particularly weather-sensitive. Residual autocorrelation (ACF(1) = 0.77) means confidence intervals understate true uncertainty: adjusting for effective sample size would widen the heat response-time CI to approximately +8–15% and the dispatch CI to +27–58%, preserving central findings but warranting caution about precise magnitudes.

Compound hot-wet events increased response time by 25.2%, approximately matching the product of individual heat (+11.7%) and precipitation (+12.1%) effects. The overall temperature×precipitation interaction surface was significant (likelihood ratio test $P < 0.001$), but this reflects departures from independence across the full exposure surface — concentrated in cold-wet conditions — rather than at the hot-wet combination specifically. The practical implication is that for summer compound events, heat and precipitation effects combine approximately multiplicatively, and planning can address each pathway somewhat independently. As compound extremes increase under climate change (Zscheischler et al., 2018), this has direct planning implications for EMS systems.

Queueing diagnostics were compatible with partially independent demand and capacity pathways. Demand elasticity remained within a narrow range (8.8–10.1%) across weather conditions, and heat-related delay persisted during lower-utilization hours, suggesting mechanisms beyond purely demand-mediated congestion — though if capacity is chronically saturated, elasticity may plateau regardless of weather, mimicking independent degradation. The contrasting delay profiles have direct intervention implications: heat delays concentrate in dispatch, where protocol-level interventions (temperature-conditional decision support) could reduce bottlenecks, whereas precipitation delays concentrate in travel, where infrastructure-level responses (pre-positioned units, routing optimization) are more appropriate.

Time-stratified analyses helped clarify lag structure. Apparent delayed heat effects were consistent with temporal mixing (afternoon exposure and nighttime events), whereas precipitation retained delayed 12–24-hour patterns across time strata. For heat, delayed effects represent ongoing burden from daytime exposures; for precipitation, true delayed effects may reflect post-storm flooding or delayed care-seeking. This distinction matters for resource planning: heat-related surge can be anticipated from temperature forecasts, whereas precipitation effects extend beyond the event itself.

Seasonal analyses revealed patterns consistent with incomplete physiological acclimatization (Balmain et al., 2018; Kenney & Munce, 2003). Within-season percentile contrasts showed elevated cardiac arrest risk in spring (RR 1.17 at 22°C versus 10°C) and summer (RR 1.25 at 31°C versus 24°C), with spring elevations occurring at moderate absolute temperatures consistent with pre-acclimatization vulnerability. Fixed-contrast analyses (30°C versus 20°C) showed a larger spring estimate (RR 1.86), but this draws from limited hours and should be interpreted with caution. These patterns suggest heat warning thresholds calibrated to summer extremes may miss clinically important spring events. Cold effects were predominantly protective, but the cardiac arrest exception during extreme winter cold underscores that EMS planning must consider both extremes (Supplementary Tables S6, S8).

Heat was associated with degraded information quality during early call processing. Non-specific classification increased across endpoints, calls coded as “caller has no patient medical information” rose during heat, severity under-triage increased for symptom-overlapping endpoints, and transport probability declined for ambiguous presentations. Cardiac arrest provided an informative contrast: severity under-triage remained null (RR 0.99), consistent with unambiguous

presentations resisting heat-related classification challenges. For stroke, under-triage reached RR 1.46 at 35°C. The reduced transport finding could reflect over-dispatch for presentations that prove heat-related at scene, system strain, or patient refusal during prolonged waits. Several observations argue against provider cognitive impairment: post-arrival scene time showed negligible heat effects, provider-level correct identification remained stable while dispatch-level non-specific classification increased (Extended Data Table 2), and severity code completeness was temperature-independent (P = 0.42). The absence of analogous classification degradation during cold reinforces this interpretation: cold symptoms are clinically distinct and do not overlap with presentations driving dispatch ambiguity. This information degradation may partly explain why dispatch — an information-processing stage — showed the largest heat-related delay (+42.6%): when callers cannot clearly describe heat-related symptoms, dispatchers require additional time to determine priority and resource assignment, identifying the caller–dispatcher interface as a specific intervention target.

Ambient temperature is not typically treated as an explicit input in Medical Priority Dispatch System workflows (Pitt, 2022; Sporer & Johnson, 2011; Sporer & Wilson, 2013). Temperature-conditioned decision support — automated prompts cueing dispatchers to probe for confusion, weakness, or altered mental status when temperature exceeds a threshold — could improve caller–dispatcher information quality without changing dispatcher roles. Because structured EMD protocols are common across EMS systems, these mechanisms may generalize beyond NYC.

Income-stratified results showed that the lowest-income quartile bore disproportionate operational burden relative to the highest. Heat effects on call volume were similar across income quartiles (heterogeneity P = 0.44), but the pre-specified Q1-Q4 response-time contrast was 2.4-fold (Q1: +20.35% versus Q4: +8.55%), where baseline call rates and response times were already higher. Because within-city temperature variation is minimal (<1°C cross-ZIP SD), the response-time difference reflects differential operational capacity rather than differential exposure. However, the pattern was not monotonic across intermediate quartiles (Q3 > Q2), consistent with borough-level infrastructure factors partially confounding the income contrast; the Q1-Q4 comparison represents a disparity between the lowest- and highest-income neighborhoods rather than a dose-response gradient, and should be interpreted as an upper bound on the income-attributable component. Nevertheless, similar relative vulnerability with unequal operational capacity suggests absolute health risk concentrates in low-income neighborhoods. Heavy precipitation increased respiratory, cardiovascular, obstetric, and pediatric calls in lower-income but not higher-income neighborhoods, consistent with housing-related exposure pathways (Krieger & Higgins, 2002). Race-stratified analysis showed non-monotonic patterns that do not parallel the income gradient (Extended Data Table 4).

Together, these findings position EMS as sentinel infrastructure for climate–health monitoring. Because EMS operates upstream of hospitals, capturing acute events at hourly resolution, it reveals system stress before emergency department or mortality data become available. The three-dimensional framework — demand surges, operational degradation, and information degradation — provides a richer and earlier signal than call volume alone, and the distinct operational signatures of different weather types (heat-driven dispatch delays versus precipitation-driven travel delays) could support real-time, targeted resource deployment during extreme weather events.

Strengths include all 911-dispatched EMS responses over 12 years, providing power for rare endpoints (cardiac arrest: ~72 calls/day) at hourly resolution, and a multi-outcome approach providing consistent evidence across demand, operational, classification, and equity analyses. Several limitations warrant consideration. EMS data miss emergencies presenting

directly to hospitals. NYC has a strong urban heat island, limiting direct generalization, though structured EMD protocols are common across systems. NLDAS-2 at ~12 km does not capture micro-scale heat island gradients, though nearest-station sensitivity analyses were consistent (within 8%). Temperature was modeled as dry-bulb; humidity and heat index were not included as primary exposures, which may underestimate physiological heat stress during high-humidity events (heat index sensitivity analyses were qualitatively consistent; Supplementary Table S5). Income stratification uses ZIP-level proxies that cluster spatially; the Q1-Q4 gradient may partly reflect borough-level operational capacity rather than income-related vulnerability. Classification analysis assumes the final EMS assessment as reference. The classification and severity analyses are exploratory and uncorrected for multiple comparisons; while pattern consistency and the null cardiac arrest contrast support the information-constraint interpretation, individual estimates should be interpreted accordingly.

As climate change increases the frequency and intensity of extreme weather events, EMS systems face compounding operational strain. These findings suggest that heat simultaneously amplifies demand, delays response, and degrades call-processing accuracy, with disproportionate burden in lower-income neighborhoods. The identification of dispatch as the most heat-sensitive response-time component — and information degradation as a plausible mechanism — points to a specific, modifiable intervention target. Prospective trials of temperature-conditioned dispatch protocols, weather-responsive resource pre-positioning, and multi-city replication are needed to translate these observational findings into operational resilience.

METHODS

Study Design and Data Sources: We conducted hourly time-series analyses of all EMS incidents in New York City from January 1, 2012, through December 31, 2023, using data from the NYC Open Data Portal. The primary analysis unit was city-hour; effect-modification analyses used separate income-stratum-hour series (one model per quartile). This study uses exclusively publicly available, de-identified data and was approved as exempt by the University of Arizona IRB. The study follows Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guidelines (von Elm et al., 2007). We designated five primary endpoints before analysis: total call volume, cardiac arrest, respiratory emergencies, total response time, and activation delay. Secondary outcomes included 16 additional clinical categories (cardiovascular, stroke, trauma, assault, mental health, poisoning/overdose, unconscious, seizures, obstetric, gastrointestinal, allergic, bleeding, hypertensive, pediatric, burns, non-specific illness) and other response time components (dispatch, travel, scene, transport). Six potential endpoints were excluded: environmental emergencies (circular with heat/cold exposure), mass casualty incidents (non-independent events), sexual assault (too sparse), child abuse (too sparse), choking (too sparse for reliable hourly DLNM estimation), and suicide attempt (too sparse).

Exposure Assessment: Hourly meteorological data (temperature, precipitation) were obtained from the North American Land Data Assimilation System Phase 2 (NLDAS-2) via Google Earth Engine, providing spatially resolved estimates at ~12 km resolution (Xia et al., 2012). ZIP code-level hourly exposures were extracted by linking ZIP code centroids to the corresponding NLDAS-2 grid cells. After excluding ZIP codes with insufficient temporal coverage (<10,000 hours, approximately 10% of the study period), 170 ZIP codes were retained for all analyses, yielding 17.9 million ZIP-hour observations (2012-2023). This approach provides finer spatial resolution than airport-based weather stations and reduces exposure measurement error. Air pollutant concentrations (fine particulate matter [$PM_{2.5}$], ozone [O_3], nitrogen dioxide [NO_2]) were obtained from the EPA Air Quality System (AQS) for sensitivity analyses. Sensitivity analyses using citywide mean exposure yielded similar results (within 8% of primary estimates).

Statistical Analysis: Analyses used four analytic units depending on the estimand: city-hour series (n = 105,192) for citywide call volume and response time DLNMs; stratum-hour series (~105,000 per quartile) for income-stratified call volume; ZIP-hour observations (n = 9.1 million with response time data) for income-stratified response time and demand elasticity; and individual incidents (n = 16.5 million) for endpoint-specific response time effects. These represent nested aggregations of the same underlying dataset, with the analytic unit chosen to match each estimand's resolution (Supplementary Table S17). Classification, severity, and transport analyses used city-hour aggregation.

Call volume models: Hourly meteorological exposures were derived from 17.9 million ZIP-hour observations (170 ZIP codes × 105,192 hours) using NLDAS-2 data. Primary demand models used a city-hour time series (n=105,192), where endpoint-specific call counts were summed across ZIP codes and exposures were population-weighted city means. Income-stratified models used separate stratum-hour time series for each income quartile (n≈105,000 observations per quartile), where calls were summed and exposures were population-weighted within quartile. We fit separate quasi-Poisson distributed lag non-linear models (DLNM) for each quartile rather than a pooled model with quartile fixed effects. NLDAS-2 at ~12 km resolution resolves approximately 5-8 distinct grid cells across NYC, so within-city hourly temperature variation is minimal (median cross-ZIP standard deviation <1°C); results were robust to unweighted averaging. Temperature was modeled using a natural cubic spline crossbasis (4 df exposure-response, 3 df lag-response) over a 24-hour lag window (Gasparrini

et al., 2010; Armstrong et al., 2019). Precipitation was modeled using a linear crossbasis with stratified lag structure (breaks at 6 and 12 hours) over the same 24-hour lag window. For income-stratified precipitation analyses (Extended Data Figure 2), we used categorical thresholds (none [<0.001 mm], light [0.001–2.5 mm], moderate [2.5–5 mm], heavy [5–10 mm], very heavy [>10 mm]) because linear precipitation terms produced tail instability in sparse strata; categorical models with temperature crossbasis adjustment were more stable. The categorical parameterization limits direct comparison with continuous citywide estimates but provides stable within-stratum contrasts. Across ZIP-hours, 83% had no measurable precipitation. Models adjusted for hour-of-day, day-of-week, seasonality spline s(doy, df=6), long-term spline s(time_idx, df=4×years), COVID-19 acute period indicator (March-June 2020), and air pollutants (PM2.5, O3, NO2). Here, time_idx is the continuous hourly index from the first to last study hour. The 4 df/year long-term trend specification was pre-specified and evaluated in sensitivity analyses. Because population is constant within each temporal stratum (city-hour or stratum-hour), it is absorbed into the intercept and does not require an explicit offset term. The model takes the form:

$$\log(E[Y_t]) = \alpha + f(\text{temperature}_t; \beta) + g(\text{precipitation}_t; \gamma) + \Sigma_j \delta_j X_{jt} + \eta_{hour} + \theta_{dow} + s(\text{doy}_t; \nu) + s(\text{time_idx}_t; \psi) + \phi \cdot \text{covid}_t$$

where $f(\cdot)$ and $g(\cdot)$ are cross-basis functions capturing non-linear and lagged exposure-response relationships, X_{jt} are air pollutant covariates (PM2.5, O3, NO2), η_{hour} and θ_{dow} are indicator variables for hour-of-day and day-of-week, $s(\text{doy}_t; \nu)$ and $s(\text{time_idx}_t; \psi)$ are natural cubic splines for seasonality (6 df) and long-term trend (4 df/year), and covid_t is an indicator for the COVID-19 acute period (March-June 2020). We evaluated residual temporal autocorrelation in model diagnostics (deviance-residual autocorrelation function [ACF] and Ljung-Box checks) for primary city-hour and representative stratified models (Supplementary Table S14). Residual ACF(1) ranged from 0.03 (cardiac arrest) to 0.77 (response time), with Ljung-Box tests significant for all models. Quasi-Poisson variance estimation accommodates overdispersion, and the DLNM crossbasis terms together with hour-of-day, day-of-week, and seasonal splines absorb most temporal structure; however, remaining autocorrelation, particularly for response time (ACF(1) = 0.77), means reported standard errors may underestimate true uncertainty. Using the approximation $n_{eff} \approx n(1-\rho)/(1+\rho)$, effective sample sizes for response time models are reduced to approximately 13% of nominal. Inflating response-time confidence intervals by the effective-N factor ($\sqrt{(1/0.13)} \approx 2.8$) would widen the heat CI from +10.4 to +13.0% to approximately +8 to +15%, preserving significance but reducing stated precision. Sparse-event endpoints (e.g., cardiac arrest, ACF(1) = 0.03) are minimally affected. Because residual autocorrelation does not bias point estimates, we interpret isolated threshold *P* values cautiously and emphasize consistency across outcomes and sensitivity analyses rather than individual significance tests.

Response time models: Response time was defined as incident_datetime to first_on_scene (call receipt to on-scene arrival). Dispatch time is incident_datetime → first_assignment, activation delay is first_assignment → first_activation (enroute), and travel time is first_activation → first_on_scene (computed as assignment-to-on-scene time minus activation delay). Scene time (arrival to departure) and transport time (departure to hospital arrival) capture the post-contact interval (contact-to-hospital transport); total incident time equals response + scene + transport (call receipt to hospital arrival). System-level operational effects were estimated with city-hour models of log(mean response time) using Gaussian regression with call-volume weights, the same 0–24-hour DLNM lag structure, and the same temporal controls as call-volume models. Income-stratified response-time effects were estimated with separate ZIP-hour DLNMs within each income quartile (one model per quartile). Temperature was modeled with DLNM crossbasis terms; precipitation was

modeled with a linear crossbasis in city-hour and ZIP-hour models. We report percent and absolute changes comparing 30°C vs 20°C (primary contrast) and 35°C vs 20°C (extreme heat), with analogous precipitation contrasts. The 35°C contrast is derived from the fitted DLNM spline surface and falls within the observed data range (maximum observed temperature: 37.9°C), though hours exceeding 35°C represent <0.2% of all observations. To distinguish direct operational effects from demand-mediated congestion, we stratified by system utilization (<80th vs ≥80th percentile hourly volume). Low-utilization hours (below the 80th percentile of hourly call volume) isolate weather effects when queueing delays are minimal, supporting interpretation as weather-related capacity degradation independent of demand surges. For endpoint-specific response time effects (Figure 2B), we fit incident-level log-linear models with concurrent linear weather terms (temperature and precipitation) and ZIP-hour clustered standard errors (n=16.5 million incidents). We used concurrent rather than DLNM specifications at the incident level because the DLNM crossbasis for precipitation was numerically unstable when fit to individual incidents (sparse high-precipitation hours yield near-singular lag matrices); ZIP-hour and city-hour DLNM results serve as the primary precipitation inference, with incident-level models providing endpoint-specific contrasts under the more conservative concurrent-exposure specification.

Response time decomposition (Figure 4; Extended Data Figure 6): ZIP-hour mean response times were modeled via DLNM with borough fixed effects to compare exposure-response shapes across components. ZIP-hour estimates capture population-level temporal patterns in system performance, while incident-level estimates provide endpoint-specific contrasts; both address complementary estimands.

Interaction and Stratification Analyses: Income-stratified analyses assign ZIP-level median household income to all incidents within that ZIP; this ecological assignment may introduce bias if within-ZIP income heterogeneity correlates with heat vulnerability, though NYC ZIP codes are relatively homogeneous in socioeconomic composition. To evaluate potential compound effects, temperature×precipitation interactions were assessed by comparing nested models in which the interaction model added cb_temp × I(precipitation>0.5 mm) to the main-effects terms (cb_temp + cb_precip). Significance was tested by analysis of deviance (F-test for quasipoisson models, chi-square fallback for gaussian models), and the relative excess risk due to interaction (RERI) was reported for call-volume interaction direction on the additive scale (super-additive vs sub-additive) (Knol and VanderWeele, 2012). Compound effects are presented for hot-dry (30°C, 0 mm), hot-wet (30°C, 5 mm), cold-dry (0°C, 0 mm), and cold-wet (0°C, 5 mm) versus reference (20°C, 0 mm). The 20°C reference was selected based on the minimum morbidity temperature commonly observed in temperate climates. Temporal stratification examined season, time-of-day, and COVID-19 period effects. Temporal stratification periods: season (winter December-February, spring March-May, summer June-August, fall September-November); time-of-day (morning rush 7:00-9:59, midday 10:00-15:59, evening rush 16:00-18:59, night 19:00-6:59); COVID-19 (pre-COVID 2012-2019, COVID March 2020-December 2021, post-COVID 2022-2023). Socioeconomic effect modification used ZIP-level income quartiles and racial composition from American Community Survey (ACS) 2016-2020 5-year estimates. Heterogeneity across income strata was assessed using Cochran's Q test and I^2 statistic on log-transformed stratum-specific RRs with inverse-variance weights, and ratio of rate ratios (RRR) with 95% CIs comparing Q1 to Q4. A composite Burden Index was calculated as (baseline call rate × heat RR × mean response time) for each income quartile, normalized to Q4 = 1.00, to summarize the joint contribution of demand, vulnerability, and operational delay to population-level strain.

Temporal Spillover Analysis: We used queueing-theory–motivated diagnostics to test whether weather-related delays are mediated by congestion: DLNM of call volume on response time with a 12-hour lag structure, decomposition into contemporaneous versus spillover effects, and stratification by capacity utilization. Technical details and M/M/c motivation are provided in the Supplementary Methods.

To distinguish demand-driven queueing from direct operational degradation, we additionally tested demand elasticity at the ZIP-hour level (n=9.1 million observations with response time data) stratified by weather type. We modeled log-transformed response time as a function of ZIP-level call volume, comparing demand elasticity (percent change in response time per additional call within a ZIP code–hour) across neutral weather (15-25°C, dry), heat (≥30°C), cold (≤0°C), heavy precipitation (>5 mm), and winter precipitation (December–February, >2 mm). Elasticity models included ZIP fixed effects and were estimated separately within each weather stratum. If weather effects on response time operated through amplified queueing, demand elasticity should be higher during adverse weather; if effects represent direct operational degradation, elasticity should remain constant. We further examined temperature × demand patterns within endpoint categories and response time components to identify condition-specific demand sensitivity.

Classification and Severity Assessment Analysis: We examined heat effects on two complementary dimensions of prehospital assessment accuracy using the same DLNM specification and city-hour aggregation as citywide call volume models. Severity code analyses were restricted to the 78.2% of incidents with both initial and final severity codes recorded; severity codes were more likely missing during high-volume periods (OR 1.03 per 10% volume increase, $P < 0.001$) but showed no association with temperature ($P = 0.42$), supporting ignorability for weather-outcome inference. First, call type classification: we compared initial call type (assigned during 911 dispatch) to final call type (determined after patient evaluation), creating city-hour counts of non-specific classifications (initial call types matching “sick,” “weak,” “faint,” “syncope,” “general,” “unknown,” “unwell,” or “malaise”) and correct condition-specific identifications per endpoint. Second, severity under-triage: we compared initial severity level code (assigned at dispatch) to final severity level code (determined after evaluation), defining under-triage as initial severity numerically exceeding final severity (higher number = lower priority; e.g., initial code 3 revised to final code 1 indicates under-triage), aggregated to city-hour counts per endpoint. Call type classification and correct identification models used quasi-Poisson DLNM with log(total endpoint incidents per hour) as offset. Severity under-triage models used log(valid severity observations per hour) as offset, restricting the denominator to incidents with both initial and final severity codes recorded. All models included temperature crossbasis (ns df=4 exposure, ns df=3 lag, 24-hour window), precipitation crossbasis (linear, strata lag with breaks at 6 and 12 hours), air pollutant adjustment (PM2.5, O3, NO2), COVID-19 acute period indicator, and temporal confounders (hour-of-day, day-of-week, seasonality 6 df, trend 4 df/year). This specification captures cumulative heat exposure effects on assessment accuracy (biologically plausible because heat exhaustion builds over hours, changing symptom presentation and information quality during 911 calls) while adjusting for temporal and environmental confounders. The parallel analysis of call type and severity provides mechanistic insight: call type misclassification reflects information constraints at the caller–dispatcher interface, while severity under-triage reflects assessment of urgency; both may be degraded by cumulative heat exposure through overlapping symptom presentations. We additionally classified initial call types into four categories (correctly identified, non-specific, heat-related, and other) and compared proportions

between heat events (>30°C) and normal conditions (10–20°C) to describe temperature-dependent shifts in classification patterns.

Transport Disposition Analysis: We examined heat effects on hospital transport probability using binomial DLNM. The INCIDENT_DISPOSITION_CODE field classifies incident outcomes: code 82 indicates hospital transport (68.9% of incidents); remaining codes indicate refusals, cancellations, gone-on-arrival, or scene releases. City-hour aggregated counts of transported versus not-transported incidents were modeled using binomial regression with logit link, temperature crossbasis (ns df=4 exposure, ns df=3 lag, 24-hour window), and the same temporal confounders as call volume models. Endpoint-specific and income-stratified transport analyses used daily aggregation for sufficient cell counts. This analysis tests the symptom-overlap mechanism: if heat is associated with reduced information specificity at the caller–dispatcher interface, leading to over-dispatch for ambiguous presentations, transport rates should decrease (scene evaluation reveals heat-related symptoms rather than medical emergency) while transport for unambiguous conditions (cardiac arrest) should remain stable.

Sensitivity Analyses: Robustness was assessed through: alternative DLNM specifications (3-5 df exposure, 2-4 df lag, 12-48 hour windows); citywide-mean versus nearest-station exposure assignment; pollutant adjustment; negative binomial models for rare endpoints (which handle overdispersion but not zero-inflation directly); and ZIP fixed effects. Primary estimates for heat effects were robust across alternative specifications: total calls RR ranged from 1.08 to 1.12 (primary: 1.10), cardiac arrest from 1.19 to 1.25 (primary: 1.22), and respiratory from 1.13 to 1.18 (primary: 1.16), with point estimates varying within ±0.02 of primary values. We also re-estimated total call volume and total response time models using heat index as the exposure; associations were qualitatively consistent, though heat index RR magnitudes were larger at the same numeric contrast because the heat index scale amplifies nonlinearly at high humidity; the two scales are not directly comparable at the same numeric threshold (Supplementary Table S5). Five primary endpoints (total calls, cardiac arrest, respiratory, total response time, activation delay) are reported at $\alpha=0.05$; the remaining 16 clinical categories are secondary endpoints with Benjamini-Hochberg false discovery rate (FDR) correction applied across these 16 secondary tests ($q<0.05$). The total analytic burden comprises 5 primary tests, 16 secondary tests for call volume, 6 response time component tests, and exploratory stratified/classification analyses. Income-stratified and classification/severity analyses are considered exploratory and are not corrected for multiple comparisons. We also assessed response-time timestamp missingness by weather (Supplementary Tables S10–S11) and re-estimated non-specific classification excluding syncope from the non-specific pattern (Supplementary Table S12).

We report relative risks (RR) for count outcomes estimated with quasi-Poisson models and odds ratios (OR) for transport disposition modeled with binomial regression. Because multiple endpoints and stratified analyses are exploratory, inference emphasizes triangulation across independent outcomes rather than isolated *P* values. Mechanistic interpretation is anchored in coherent patterns across classification, severity, transport, and response-time localization, with cardiac arrest severity under-triage specified a priori as a negative-control contrast for diagnostic ambiguity.

REFERENCES

Alessandrini, E., Zauli Sajani, S., Scotto, F., et al. (2011). Emergency ambulance dispatches and apparent temperature: a time-series analysis in Emilia-Romagna, Italy. *Environmental Research*, 111, 1192-1200.

Armstrong, B., Sera, F., Vicedo-Cabrera, A. M., et al. (2019). The role of humidity in associations of high temperature with mortality: a multicountry, multicity study. *Environmental Health Perspectives*, 127, 97007.

Balmain, B. N., Sabapathy, S., Louis, M., Morris, N. R., & Langhammer, B. (2018). Aging and thermoregulatory control: the clinical implications of exercising under heat stress in older individuals. *BioMed Research International*, 2018, 8306154.

Cheng, J., Xu, Z., Bambrick, H., et al. (2016). The burden of extreme heat and heatwave on emergency ambulance dispatches: a time-series study in Huainan, China. *Science of the Total Environment*, 571, 27-33.

Cui, Y., Ai, S., Liu, Y., et al. (2020). Hourly associations between ambient temperature and emergency ambulance calls in one central Chinese city: call for an immediate emergency plan. *Science of the Total Environment*, 711, 135046.

Gasparrini, A., Armstrong, B., & Kenward, M. G. (2010). Distributed lag non-linear models. *Statistics in Medicine*, 29, 2224-2234.

Green, L. V., Soares, J., Giglio, J., & Green, R. (2006). Using queueing theory to increase the effectiveness of emergency department provider staffing. *Academic Emergency Medicine*, 13, 61-68.

Hu, X., Barnes, S., & Golden, B. (2018). Applying queueing theory to the study of emergency department operations: a survey. *International Transactions in Operational Research*, 25, 21-49.

Kenney, W. L., & Munce, T. A. (2003). Invited review: aging and human temperature regulation. *Journal of Applied Physiology*, 95, 2598-2603.

Knol, M. J., & VanderWeele, T. J. (2012). Recommendations for presenting analyses of effect modification and interaction. *International Journal of Epidemiology*, 41, 514-520.

Kotani, K., Ueda, K., Seposo, X., et al. (2018). Effects of high ambient temperature on ambulance dispatches in different age groups in Fukuoka, Japan. *Global Health Action*, 11, 1437882.

Krieger, J., & Higgins, D. L. (2002). Housing and health: time again for public health action. *American Journal of Public Health*, 92, 758-768.

Lai, P. H., Lancet, E. A., Weiden, M. D., et al. (2020). Characteristics associated with out-of-hospital cardiac arrests and resuscitations during the novel coronavirus disease 2019 pandemic in New York City. *JAMA Cardiology*, 5, 1154-1163.

Ma, W., Zeng, W., Zhou, M., et al. (2015). The short-term effect of heat waves on mortality and its modifiers in China: an analysis from 66 communities. *Environment International*, 75, 103-109.

Morgan, D. P., Muscatello, D., Travaglia, J., & Hayen, A. (2020). Retrospective observational cohort study of out-of-hospital cardiac arrest outcomes in Tasmania 2010-2014. *Emergency Medicine Australasia*, 32, 631-637.

Perkins-Kirkpatrick, S. E., & Lewis, S. C. (2020). Increasing trends in regional heatwaves. *Nature Communications*, 11, 3357.

Pitt, I. L. (2022). The system-wide effects of dispatch, response and operational performance on emergency medical services during Covid-19. *Humanities and Social Sciences Communications*, 9(1), 412.

Riva, G., & Hollenberg, J. (2021). Different forms of bystander cardiopulmonary resuscitation in out-of-hospital cardiac arrest. *Journal of Internal Medicine*, 290, 57-72.

Ruiz de Gauna, S., Irusta, U., Ruiz, J., et al. (2014). Rhythm analysis during cardiopulmonary resuscitation: past, present, and future. *BioMed Research International*, 2014, 386010.

Seong, K., Jiao, J., & Mandalapu, A. (2023). Hourly associations between heat index and heat-related emergency medical service (EMS) calls in Austin-Travis County, Texas. *International Journal of Environmental Research and Public Health*, 20, 6853.

Sporer, K. A., & Johnson, N. J. (2011). Detailed analysis of prehospital interventions in Medical Priority Dispatch System determinants. *Western Journal of Emergency Medicine*, 12(1), 19-29.

Sporer, K. A., & Wilson, K. G. (2013). How well do emergency medical dispatch codes predict prehospital medication administration in a diverse urban community? *Journal of Emergency Medicine*, 44(2), 413-422.

Turner, L. R., Connell, D., & Tong, S. (2013). The effect of heat waves on ambulance attendances in Brisbane, Australia. *Prehospital and Disaster Medicine*, 28, 482-487.

von Elm, E., Altman, D. G., Egger, M., et al. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *The Lancet*, 370, 1453-1457.

Wang, Y. C., & Lin, Y. K. (2020). Ambulance services associated with extreme temperatures and fine particles in a subtropical island. *Scientific Reports*, 10, 2855.

Wataki, H., Seposo, X. T., Uttajug, A., et al. (2026). Impact of snowfall on ambulance response times to out-of-hospital cardiac arrest – nationally-representative population-based registry analysis, 2009–2019. *Circulation Reports*, advance online publication. doi:10.1253/circrep.CR-25-0129.

Wazneh, H., Arain, M. A., & Coulibaly, P. (2020). Evaluating the dependence between temperature and precipitation to better estimate the risks of concurrent extreme weather events. *Advances in Meteorology*, 2020, 8763631.

Xia, Y., Mitchell, K., Ek, M., et al. (2012). Continental-scale water and energy flux analysis and validation for the North American Land Data Assimilation System project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. *Journal of Geophysical Research: Atmospheres*, 117, D03109.

Xu, Z., Watzek, J. T., Phung, D., et al. (2023). Heat, heatwaves, and ambulance service use: a systematic review and meta-analysis of epidemiological evidence. *International Journal of Biometeorology*, 67, 1523-1542.

Yuan, L., Madaniyazi, L., Vicedo-Cabrera, A. M., et al. (2023). A nationwide comparative analysis of temperature-related mortality and morbidity in Japan. *Environmental Health Perspectives*, 131, 127008.

Zhan, Z. Y., Yu, Y. M., Chen, T. T., Xu, L. J., An, S. L., & Ou, C. Q. (2020). Effects of hourly precipitation and temperature on ambulance response time. *Environmental Research*, 181, 108946.

Zscheischler, J., Westra, S., van den Hurk, B. J. J. M., et al. (2018). Future climate risk from compound events. *Nature Climate Change*, 8, 469-477.

Ethics Statement: This study uses exclusively publicly available, de-identified data from the NYC Open Data Portal. The study protocol was reviewed and approved as exempt by the University of Arizona Institutional Review Board.

Community Engagement and Patient/Public Involvement: This study analyzed publicly available, de-identified administrative data with no direct participant contact; formal patient/public involvement was therefore not applicable. Endpoint selection and interpretation of operational findings were informed by co-investigators with EMS domain expertise (NCM, JAR). Study findings will be disseminated to NYC EMS operational leadership and community health organizations in the neighborhoods identified as bearing disproportionate weather-related burden, with the goal of informing dispatch protocol adaptation and resource allocation.

Keywords: emergency medical services, climate change, heat, distributed lag non-linear model, health equity, prehospital care, prehospital monitoring

Funding: National Institute of Environmental Health Sciences (R01ES034740)

Competing Interests: The authors declare no competing interests.

Data Availability: EMS incident data are available from the NYC Open Data Portal (https://data.cityofnewyork.us). Meteorological data are from the North American Land Data Assimilation System Phase 2 (NLDAS-2) via Google Earth Engine.

Code Availability: Analysis code will be deposited in a public repository upon publication.

Software: All analyses were conducted in R (version 4.5.1) using the dlnm (version 2.4.10), data.table, splines, and MASS packages. Figures were generated using ggplot2.

Author Contributions: CCL conceptualized the study, designed the analysis, and wrote the manuscript. JAR contributed to study design and data interpretation. HK contributed to study design statistical methodology. RW contributed to spatial analysis methods. XL contributed to environmental exposure assessment. LG contributed to epidemiological design and data interpretation. CC contributed to study design and data interpretation. NCM contributed to study design and provided EMS domain expertise and contributed to data interpretation. All authors reviewed and approved the final manuscript.

Acknowledgments: The authors acknowledge the NYC Open Data Portal for making EMS incident data publicly available.

DISPLAY ITEMS

Table 1. Study Characteristics by ZIP Code Income Quartile, New York City Emergency Medical Services, 2012-2023

Characteristic	Q1 (Lowest)	Q2	Q3	Q4 (Highest)	Total
Neighborhood Characteristics					
N ZIP codes	43	42	42	42	169[e]
Mean household income	$39,443	$60,725	$78,224	$119,308	–
Poverty rate, %	28.8	16.2	10.7	8.7	16.2
Non-white population, %	74.2	61.6	52.0	31.9	55.1
EMS Incidents					
Total incidents, N	6,834,024	4,133,887	2,274,160	3,244,222	16,486,293
Incidents (% of total)	41.5%	25.1%	13.8%	19.7%	100%
Operational Times, seconds					
Response time, mean[a]	615	538	529	546	570
Dispatch time, mean[b]	92	62	47	75	75
Activation time, mean[c]	27	24	22	24	25
Travel time, mean[d]	491	451	461	444	468

Abbreviations: EMS, Emergency Medical Services; Q, quartile.

[a]Response time: call receipt to patient contact/on-scene arrival (dispatch + activation + travel; used in weather-effect analyses). [b]Dispatch time: call receipt to unit assignment. [c]Activation time: unit assignment to en route. [d]Travel time: en route to on-scene arrival. Scene time (arrival to departure) and transport time (departure to hospital arrival) are reported in response-time component analyses; total incident time equals response + scene + transport. Values are call-weighted means from ZIP-hour aggregated data; response time is computed directly from timestamps and component means are independently weighted, so components may not sum exactly to total. The non-monotonic dispatch time across quartiles (Q3 < Q2 < Q4) reflects geographic and operational variation that does not follow the income gradient. [e]169 ZIP codes had valid income quartile assignment; 1 ZIP code lacked ACS income data.

Note: Income quartiles based on ZIP code median household income from American Community Survey (ACS) 2016-2020 5-year estimates. Q1 represents lowest income neighborhoods (mean $39,443), Q4 represents highest income neighborhoods (mean $119,308). N = 16,486,293 incidents across 169 ZIP codes with valid income quartile assignment. Non-white percentages are unweighted ZIP-level averages; the population-weighted citywide value is 57.3%.

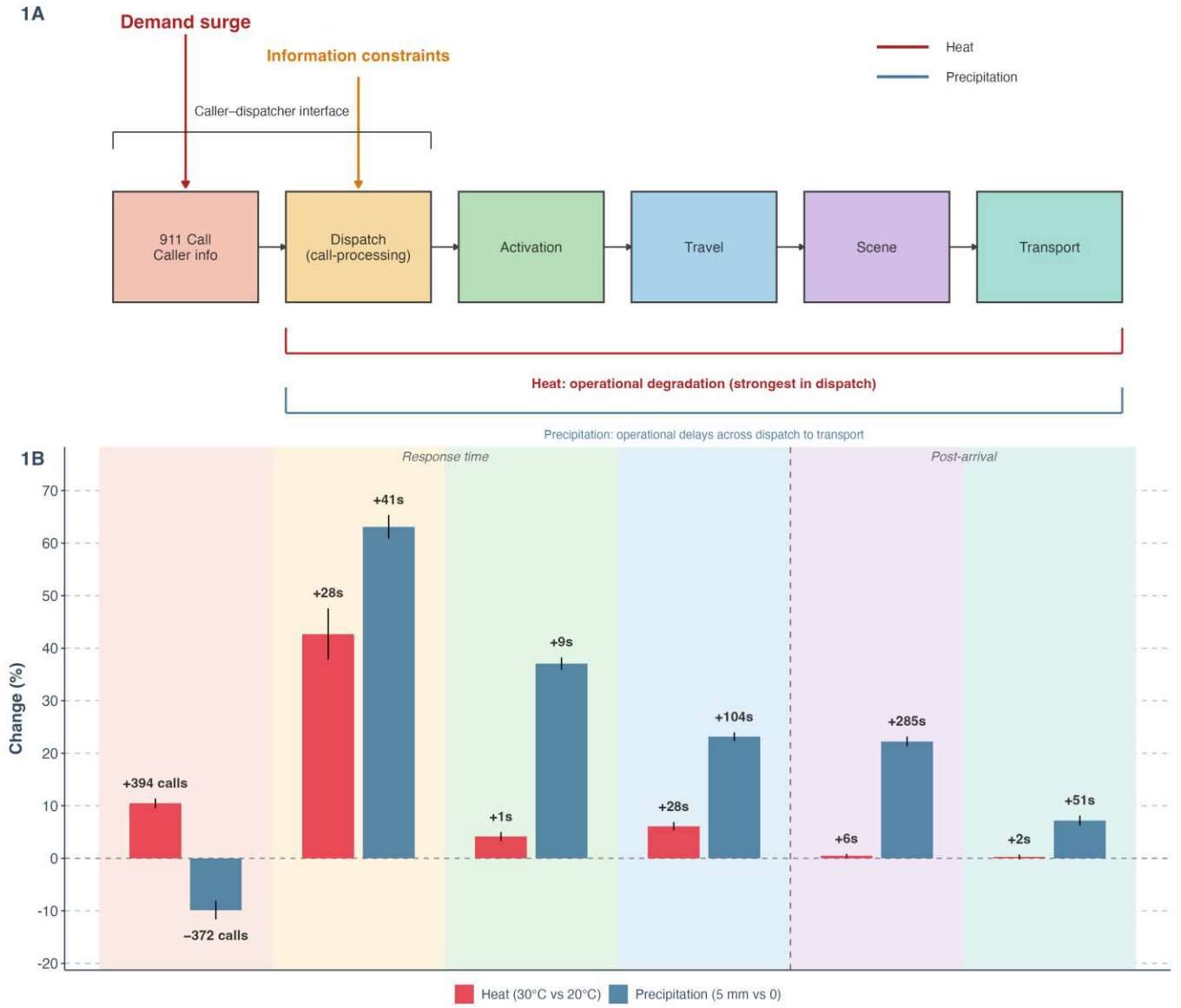


Figure 1. Conceptual framework and quantitative pipeline effects. (A) Schematic linking extreme weather to emergency medical services (EMS) stress. Emergency responses progress from the 911 call through dispatch (call-processing), activation, travel, scene care, and transport. Extreme heat is associated with a dual burden: increased demand (higher call volume) and degraded response capacity. Information constraints at the caller-dispatcher interface may increase diagnostic ambiguity for clinically overlapping conditions, while operational degradation extends across dispatch through transport. Arrows indicate hypothesized pathways consistent with empirical findings; the figure is conceptual and not intended to imply causal attribution. (B) Percent change in each pipeline component under heat (30°C vs 20°C; city-hour DLNM) and heavy precipitation (5 mm vs 0; ZIP-hour DLNM), with absolute changes annotated (seconds for time components; calls per day for 911 call volume). Heat increased call volume by 10% (+394 calls/day) and dispatch time by 42.6% (+28 seconds), while precipitation reduced call volume by 9.9% (−372 calls/day) but increased dispatch time by 63.1% (+41 seconds). Effects attenuate downstream through the pipeline for both exposures. Percent changes are relative to model-predicted values at reference conditions (20°C, 0 mm precipitation). Important: heat estimates use city-hour DLNM and precipitation estimates use ZIP-hour DLNM; these derive from different analytic units and magnitudes are not directly comparable across exposures.

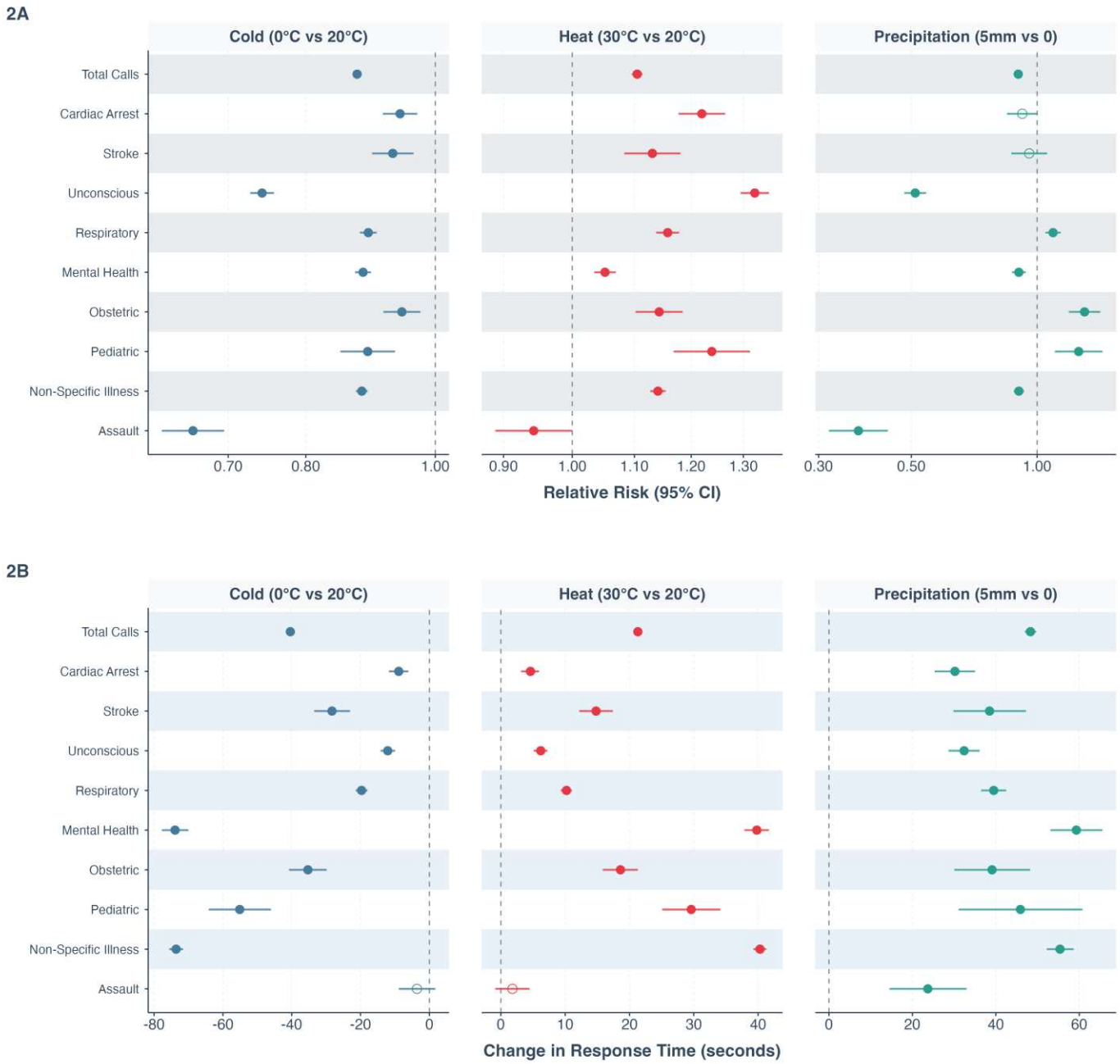
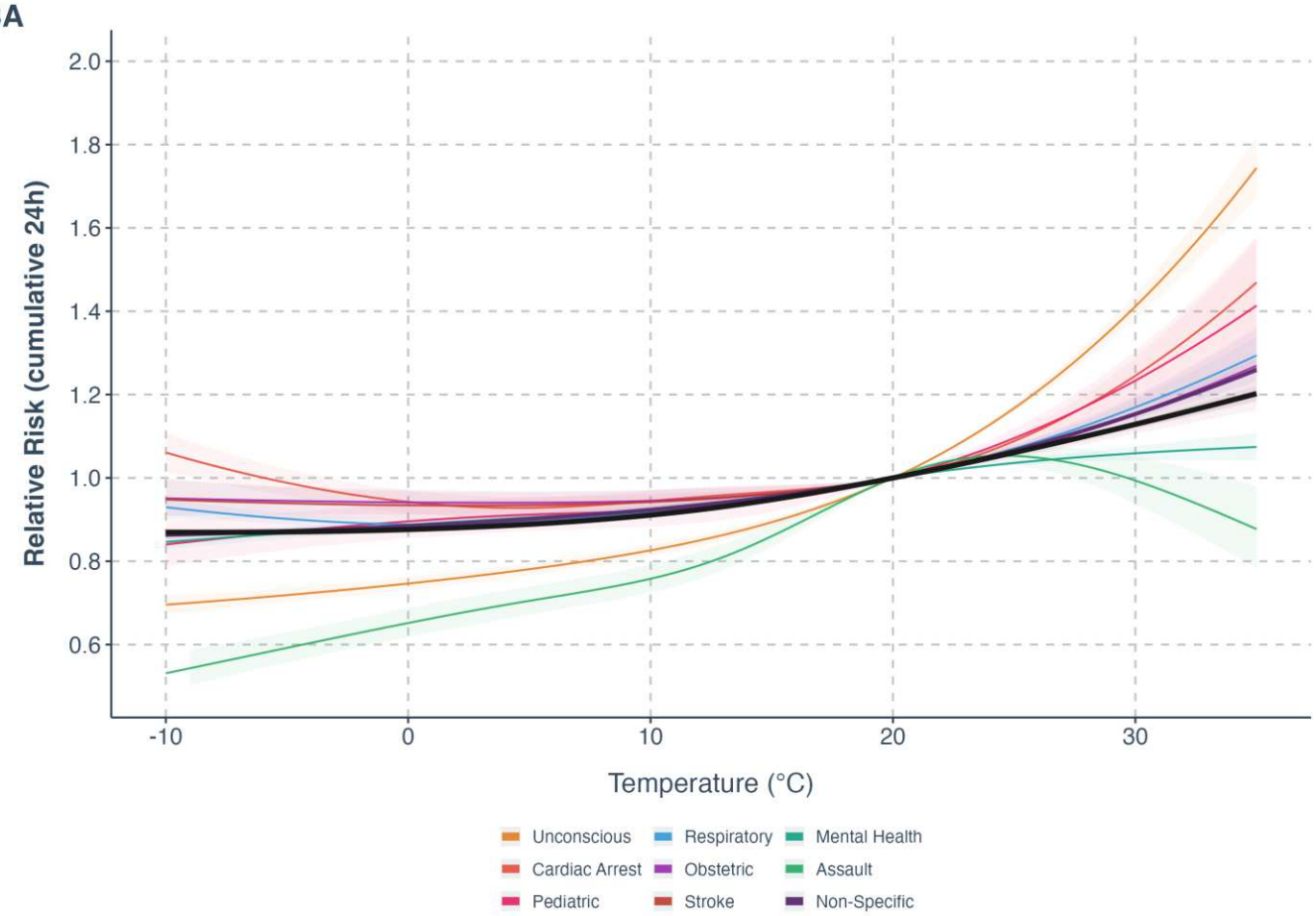


Figure 2. Weather effects on EMS call volume and response time for 10 core clinical endpoints. Forest plots showing cold (0°C vs 20°C), heat (30°C vs 20°C), and heavy precipitation (5 mm vs 0 mm) effects on (A) EMS call volume (cumulative 24-hour DLNM) and (B) response time (concurrent incident-level model). Panel A: Cumulative rate ratios from city-hour distributed lag non-linear models (n=105,192 city-hours) with 24-hour lag structure; exposures derived from 17.9 million ZIP-hour observations. Panel B: Percent change in response time (call receipt to patient contact) from incident-level log-linear regression with concurrent linear weather exposure (no DLNM); n=16.5 million incidents. Both adjusted for hour, day-of-week, trend, and seasonality. Panels A and B use different statistical models and estimands (cumulative DLNM rate ratios vs concurrent per-call percent change); magnitudes are not directly comparable across panels. Panel A uses quasi-Poisson sandwich variance; Panel B uses robust standard errors clustered at the ZIP-hour level. Horizontal lines show 95% confidence intervals. Dashed vertical lines indicate null effect (RR=1.0 for call volume, 0 seconds for response time). Full 19-endpoint results in Extended Data Figure 5. New York City, 2012-2023.

3A


Relative Risk (cumulative 24h)
2.0
1.8
1.6
1.4
1.2
1.0
0.8
0.6
-10
0
10
20
30
Temperature (°C)
Unconscious
Cardiac Arrest
Pediatric
Respiratory
Obstetric
Stroke
Mental Health
Assault
Non-Specific

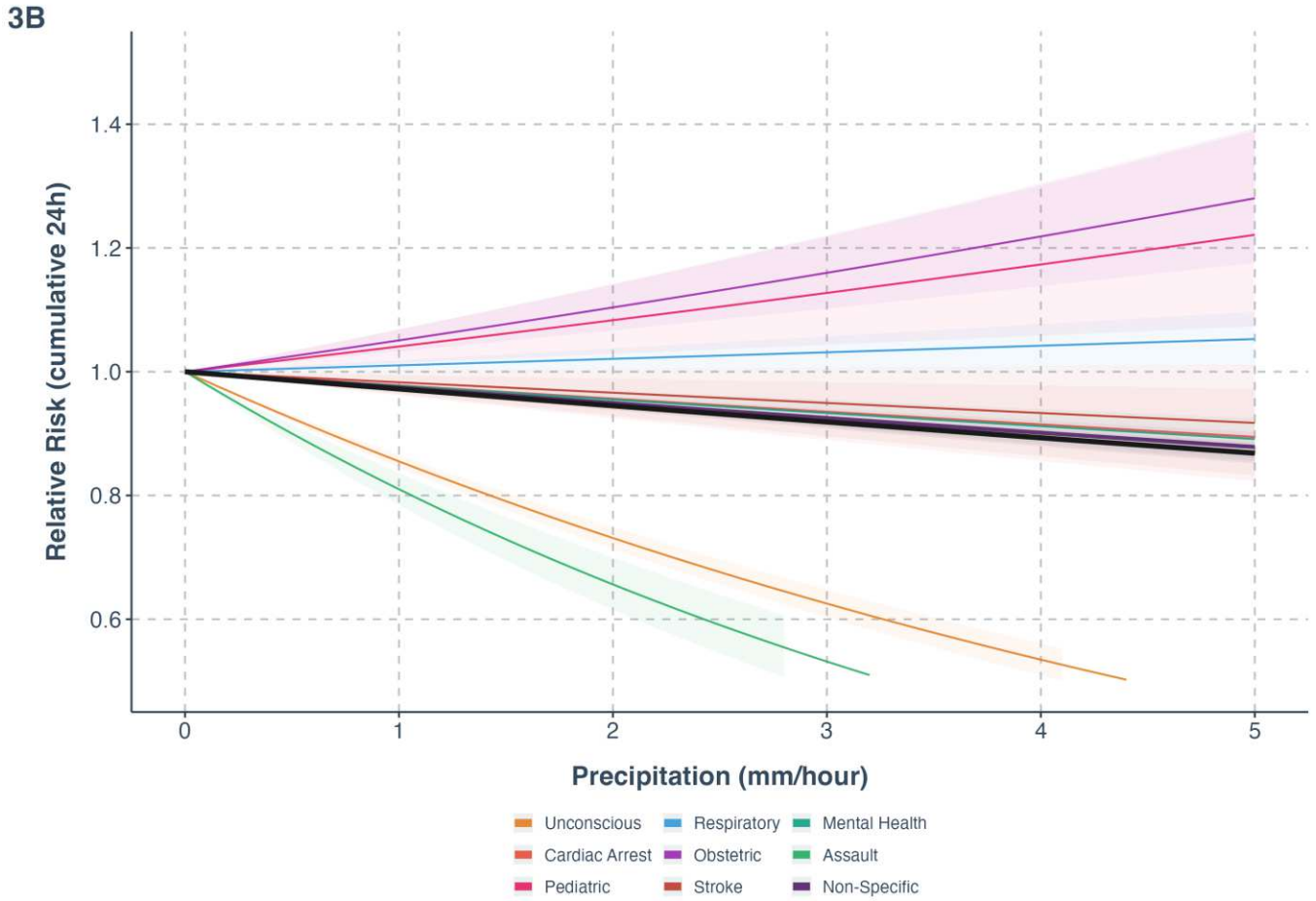


Figure 3. Temperature and precipitation dose-response relationships for 10 core clinical endpoints. Exposure-response curves showing cumulative relative risk (95% CI) for EMS call volume across (A) temperature and (B) precipitation ranges. Total Calls shown in black (thicker line); other endpoints in distinct colors. Reference values: 20°C for temperature, 0 mm for precipitation. Shaded areas represent 95% confidence bands. Curves derived from distributed lag non-linear models with natural cubic splines (4 df for temperature exposure-response, 3 df for lag-response; linear crossbasis for precipitation with 0 mm reference). Models adjusted for hour, day-of-week, trend, and seasonality. City-hour DLNM (n=105,192 city-hours); exposures derived from 17.9 million ZIP-hour observations. Full 19-endpoint results in Extended Data Figure 1. New York City, 2012-2023.

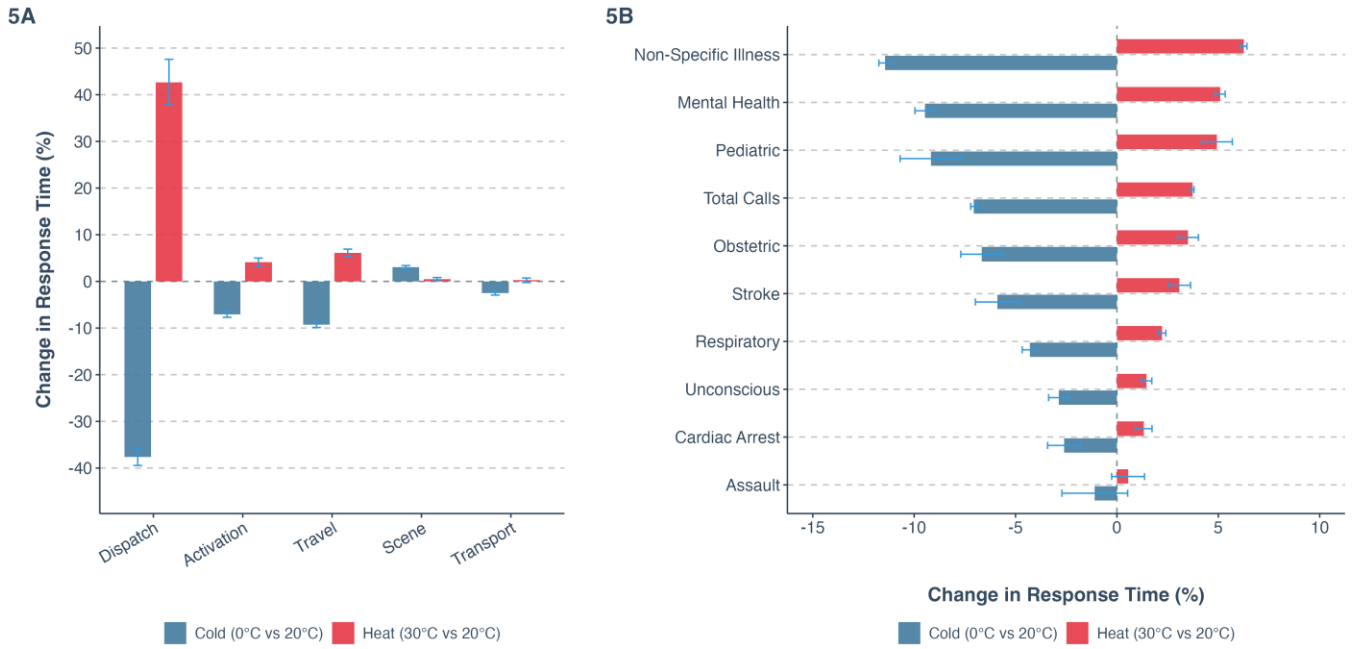


Figure 4. Response time decomposition by component and clinical outcome. (A) Percent change in response time components comparing heat (30°C) and cold (0°C) to 20°C reference. Heat was associated with increased dispatch time (+42.6%), activation delay (+4.1%), and travel time (+6.1%), with smaller effects on scene time and transport time. Cold (0°C) was associated with reductions in dispatch (-37.7%), travel (-9.2%), and activation (-7.1%), but increased scene time (+3.1%). (B) Response time changes by clinical outcome for 10 core endpoints. Heat effects on total response time vary by endpoint; full 19-endpoint results in Supplementary Figure S6. Panel A uses city-hour DLNM (n=105,192 city-hours); Panel B uses incident-level log-linear models (n=16.5 million incidents). Models adjusted for hour, day-of-week, seasonality, and trend. Error bars show 95% confidence intervals. New York City, 2012-2023.

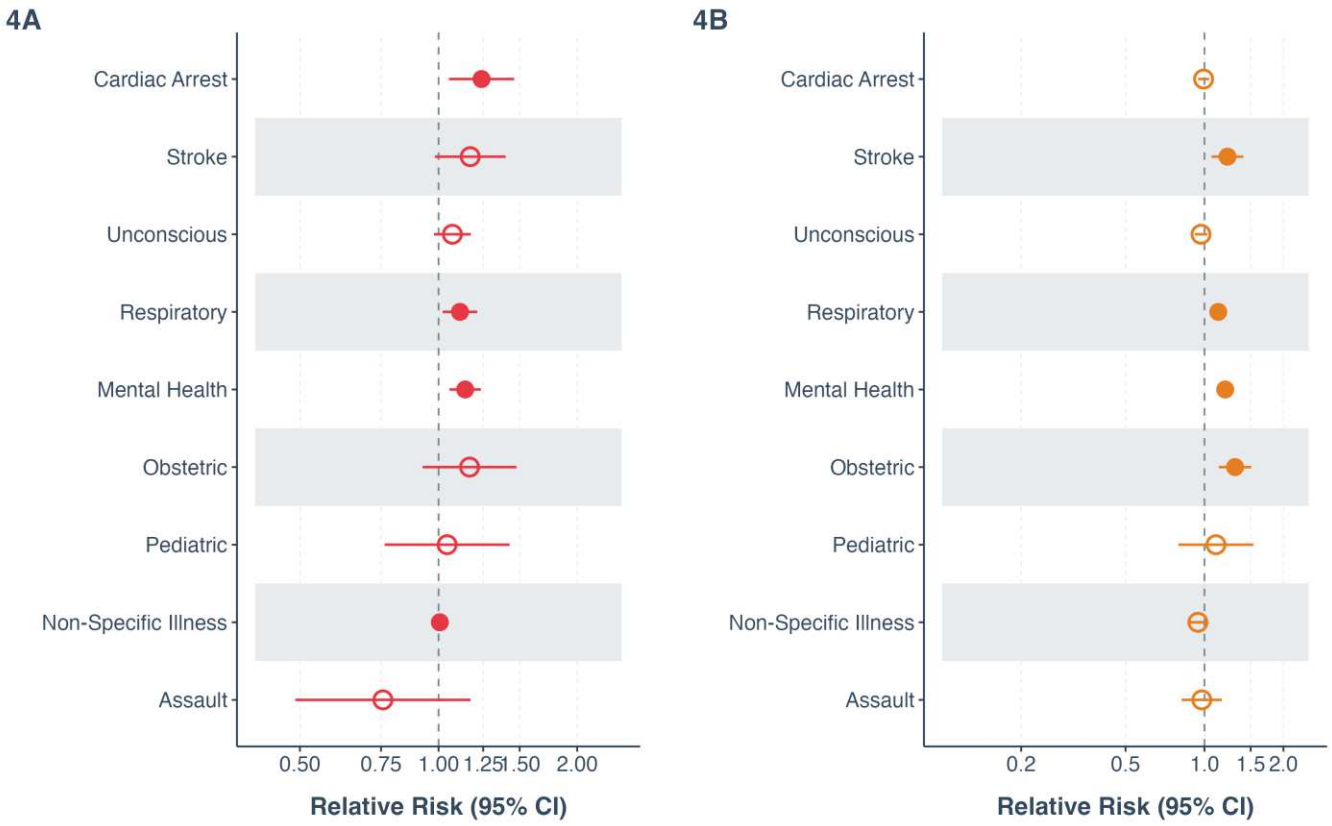


Figure 5. Heat effects on EMS classification patterns. Forest plots showing relative risk (30°C vs 20°C reference) for (A) non-specific classification and (B) severity under-triage by clinical endpoint. Among core endpoints, non-specific classification increased for cardiac arrest (RR 1.24) and mental health (1.14). Severity under-triage increased for obstetric (RR 1.31), stroke (1.22), and mental health (1.20). Cardiac arrest severity assessment remained unaffected (RR 0.99), consistent with unambiguous presentations being resistant to heat-related classification challenges. Filled circles indicate statistically significant effects (95% CI excludes 1.0); open circles indicate non-significant effects. Full results for all 19 endpoints in Supplementary Figure S5. Cumulative 24-hour relative risk from quasi-Poisson DLNM adjusted for hour, day-of-week, seasonality, and trend. City-hour aggregated models (n=105,192 city-hours); counts derived from 16.5 million incidents (severity analyses restricted to 78.2% with valid initial and final severity codes). New York City, 2012-2023.

EXTENDED DATA

Extended Data Table 1. Temperature-Precipitation Compound Effects on EMS Response Time

Condition	Response Time Change	95% CI	*P* value
Heat alone (30°C, 0 mm)	+11.7% (+65.4s)	+10.4 to +13.0	<0.001
Rain alone (20°C, 5 mm)	+12.1% (+67.5s)	+9.1 to +15.1	<0.001
Expected additive (arithmetic)	+23.8%	–	–
Compound (30°C, 5 mm)	+25.2% (+140.8s)	+21.5 to +28.9	<0.001
Interaction (LRT)	–	–	1.9×10^{-7}

Note: The compound effect (+25.2%) is consistent with the multiplicative combination of individual effects (1.117 × 1.121 = 1.252) and exceeds the arithmetic sum (+23.8%) by 1.4 percentage points. The overall temperature×precipitation interaction surface is significant (LRT P* < 0.001), indicating joint effects differ from independent main effects across the exposure range. Interaction assessed by likelihood ratio test comparing DLNM with and without temperature×precipitation(binary) crossbasis interaction. Response time defined as call receipt to patient contact. Log-linear Gaussian DLNM with call-volume weights, adjusted for seasonality (ns, 6 df), long-term trend (ns, 4 df/year), hour, day-of-week, $PM_{2.5}$, O_3, NO_2. 798 hours with T>30°C and measurable precipitation. n=105,192 city-hour observations, New York City, 2012-2023.*

Extended Data Table 2. DLNM-Adjusted Heat Effects on Classification and Severity Assessment

Endpoint	Non-specific RR (30°C)	Non-specific RR (35°C)	Correct ID RR (30°C)	Correct ID RR (35°C)	Under-Triage RR (30°C)	Under-Triage RR (35°C)
Stroke	1.17 (0.98-1.40)	1.34 (0.95-1.89)	1.03 (1.01-1.05)	1.05 (1.01-1.10)	1.22 (1.06-1.41)	1.46 (1.11-1.92)
Mental Health	1.14 (1.06-1.24)	1.30 (1.11-1.51)	0.96 (0.95-0.97)	0.93 (0.91-0.94)	1.20 (1.13-1.28)	1.40 (1.24-1.58)
Cardiovascular	1.12 (1.05-1.19)	1.21 (1.08-1.37)	0.99 (0.98-1.00)	0.98 (0.96-1.00)	1.10 (1.05-1.15)	1.16 (1.06-1.27)
Respiratory	1.11 (1.02-1.21)	1.16 (0.98-1.37)	0.99 (0.99-1.00)	0.98 (0.97-0.99)	1.13 (1.07-1.19)	1.19 (1.07-1.32)
Seizures	1.25 (1.08-1.45)	1.48 (1.11-1.98)	0.99 (0.95-1.03)	0.98 (0.91-1.06)	1.19 (1.09-1.30)	1.38 (1.16-1.64)
Cardiac Arrest	1.24 (1.05-1.46)	1.44 (1.05-1.98)	1.03 (1.01-1.05)	1.05 (1.02-1.09)	0.99 (0.95-1.04)	0.98 (0.90-1.08)
Obstetric	1.17 (0.92-1.48)	1.34 (0.85-2.12)	1.01 (0.99-1.03)	1.01 (0.97-1.06)	1.31 (1.13-1.51)	1.52 (1.15-2.01)
Trauma	1.12 (1.02-1.22)	1.26 (1.06-1.50)	1.00 (0.99-1.00)	0.99 (0.98-1.00)	1.08 (1.03-1.13)	1.12 (1.02-1.22)
Assault	0.76 (0.49-1.17)	0.61 (0.26-1.44)	1.00 (0.97-1.04)	0.96 (0.89-1.03)	0.98 (0.82-1.17)	0.94 (0.66-1.33)

Interpretation guide: Non-specific RR >1 = increased vague classification (worse); Correct ID RR <1 = reduced condition-specific identification (worse); Under-Triage RR >1 = increased severity underestimation (worse). Values are cumulative 24-hour relative risks (95% CI) from quasi-Poisson DLNM comparing indicated temperature to 20°C reference. Call type models use offset log(total endpoint incidents/hour); severity under-triage models use offset log(valid severity observations/hour). Models adjusted for hour, day-of-week, seasonality (6 df), long-term trend (4 df/year), and air pollutants ($PM_{2.5}$, O_3, NO_2). Population-weighted city-hour weather from NLDAS-2. Cardiac arrest shows null severity under-triage at both thresholds (RR 0.99 and 0.98), consistent with the symptom-overlap mechanism: unambiguous presentations are not under-triaged regardless of caller difficulty describing symptoms. City-hour aggregated models (n=105,192 city-hours); counts derived from 16.5 million incidents. New York City, 2012-2023.

Extended Data Table 3. Weather Effects on EMS Call Volume and Response Time by Neighborhood Income Quartile

Income Quartile	Baseline Call Rate	Baseline RT	Heat Call RR	Heat RT Δ	Cold Call RR	Cold RT Δ	Precip Call RR	Precip RT Δ	Burden Index
Q1 (Lowest)	57.9/100k/d	615 s	1.10 (1.09-1.11)	+20.35%	0.87 (0.86-0.88)	−10.43%	0.91 (0.89-0.93)	+35.10%	1.39×
Q2	40.7/100k/d	538 s	1.11 (1.10-1.12)	+10.22%	0.88 (0.87-0.89)	−8.38%	0.90 (0.87-0.94)	+20.00%	0.87×
Q3	29.1/100k/d	529 s	1.10 (1.09-1.12)	+14.30%	0.89 (0.87-0.90)	−10.94%	0.90 (0.87-0.94)	+13.73%	0.60×
Q4 (Highest)	46.8/100k/d	546 s	1.10 (1.09-1.12)	+8.55%	0.88 (0.87-0.89)	−12.03%	0.89 (0.85-0.92)	+6.15%	1.00 (ref)

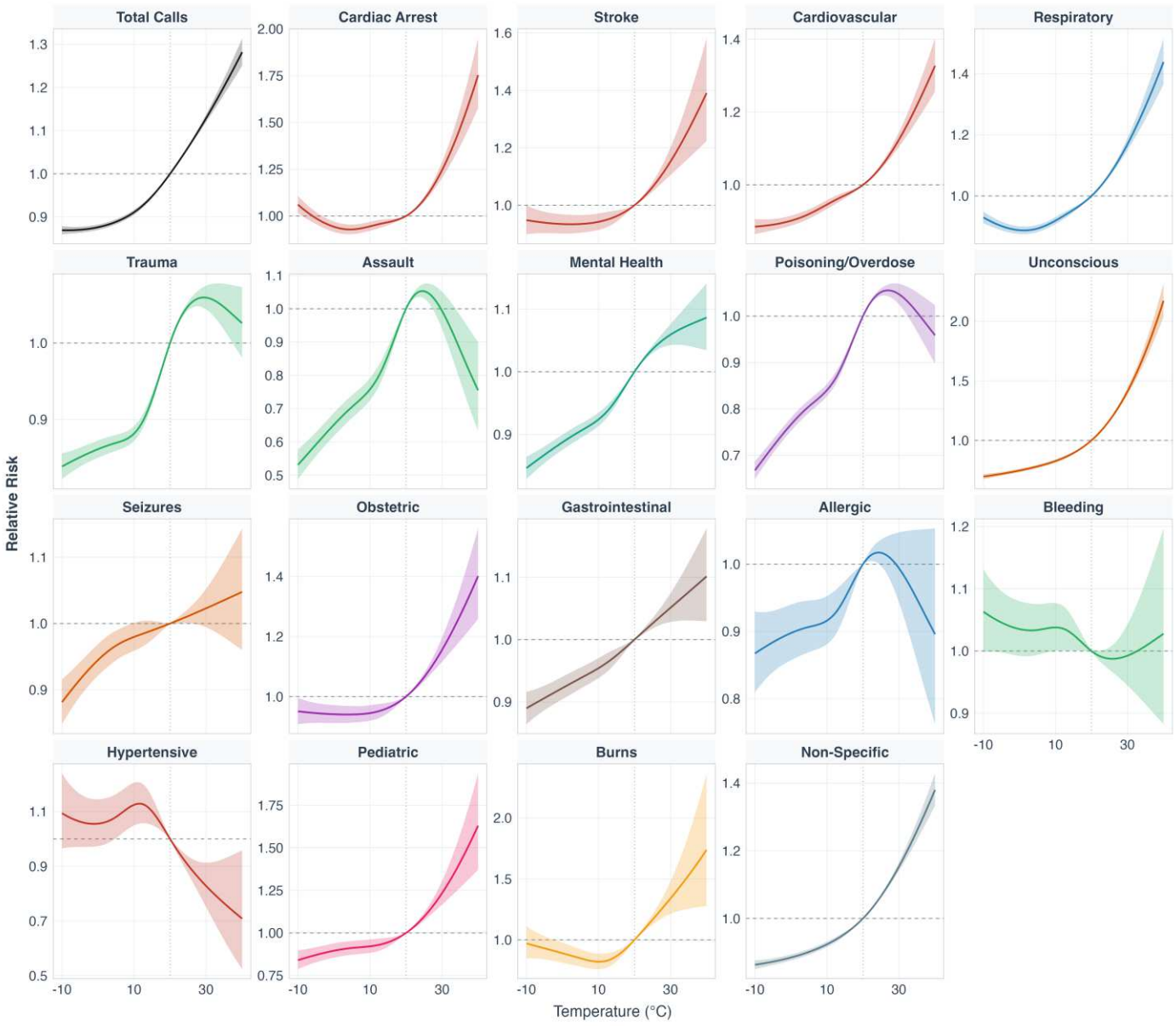
Abbreviations: RR, relative risk; RT, response time; Δ, change.

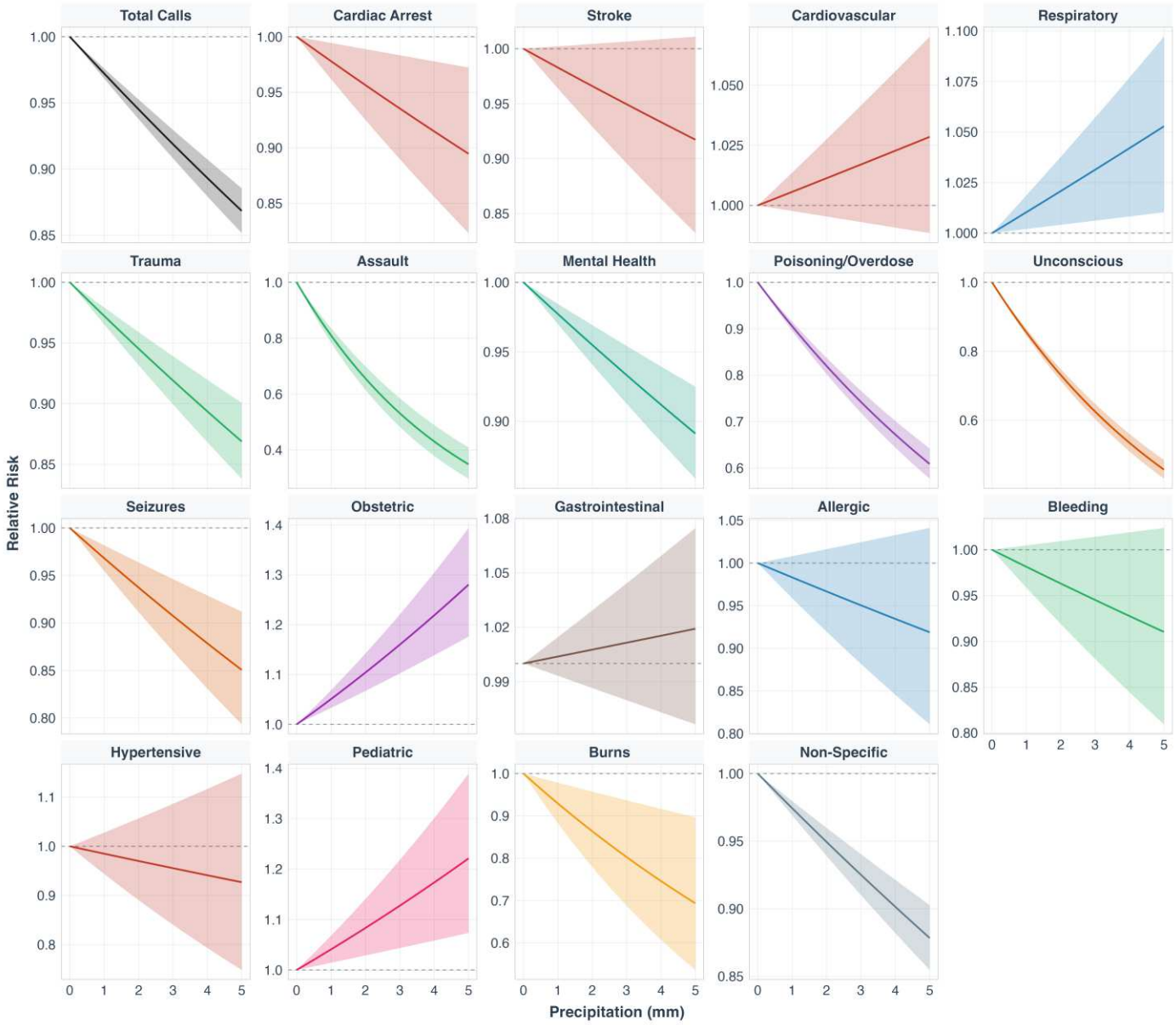
Note: Heat = 30°C vs 20°C; Cold = 0°C vs 20°C; Precip = 5 mm vs 0 mm. Burden Index calculated as (baseline call rate × heat RR × mean response time) for each quartile, normalized to Q4 = 1.00. Call volume RR from stratum-hour quasi-Poisson DLNM; RT change from ZIP-hour Gaussian DLNM. Call volume effects are uniform across income quartiles for both heat (~10-11%; Cochran's Q=2.69, P* = 0.44) and cold (~12-13%). Response-time effects show income gradients for both heat (Q1 +20.35% vs Q4 +8.55%; ratio 2.4) and precipitation (Q1 +35.10% vs Q4 +6.15%; ratio 5.7), with the precipitation gradient substantially steeper. Cold RT reductions show no clear income gradient (-8% to -12% across quartiles). Response-time percent changes are point estimates from separate within-quartile ZIP-hour DLNMs; between-quartile differences should be interpreted as descriptive comparisons rather than formal statistical tests.*

Extended Data Table 4. Race Composition-Stratified Precipitation Effects on EMS Call Volume (5 mm vs 0 mm)

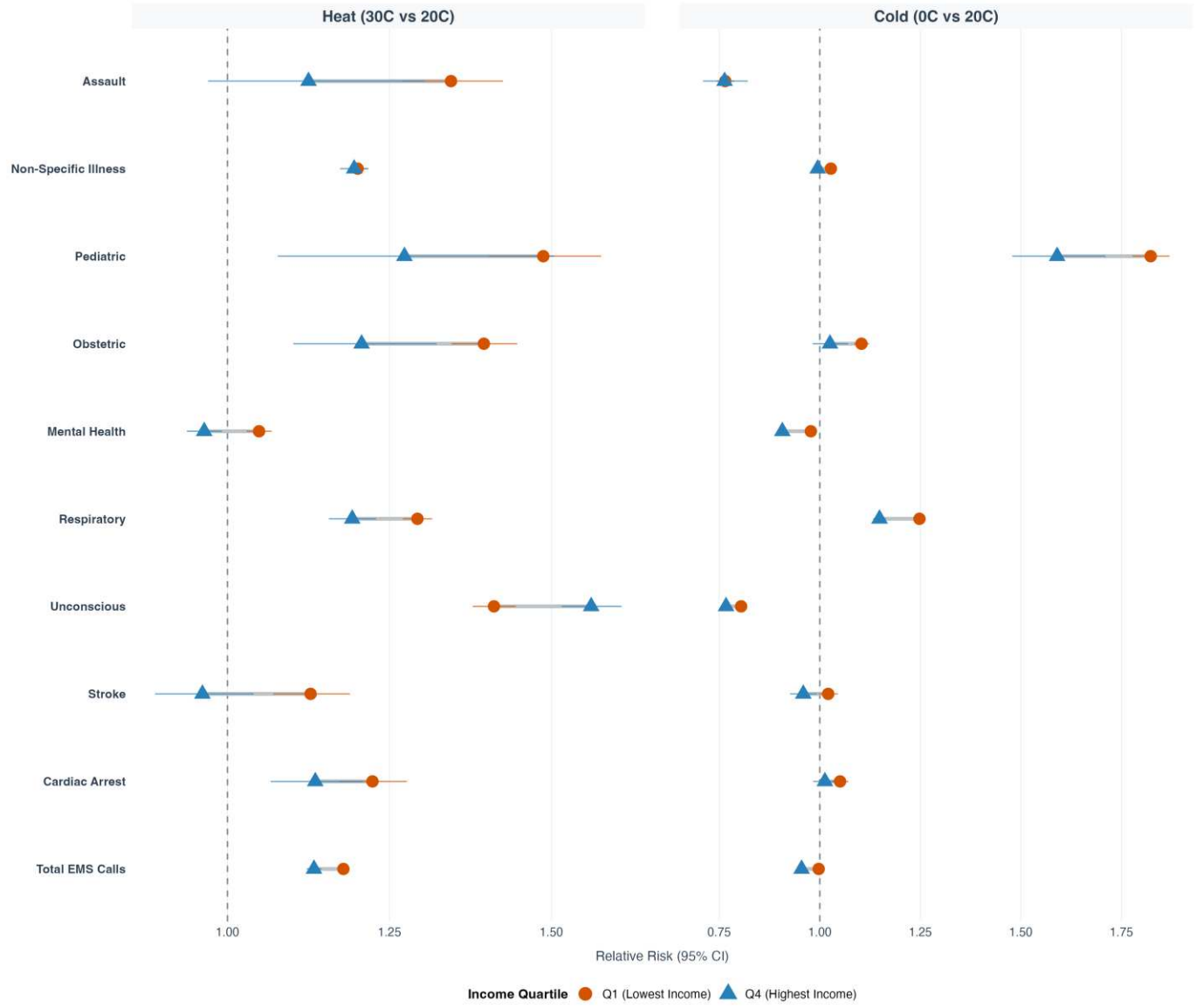
Endpoint	Q1 (22% non-White)	Q2 (44% non-White)	Q3 (67% non-White)	Q4 (88% non-White)
Total Calls	0.89 (0.86-0.92)*	0.87 (0.84-0.90)*	0.91 (0.89-0.93)*	0.93 (0.91-0.95)*
Cardiac Arrest	0.98 (0.81-1.19)	0.83 (0.68-1.01)	0.90 (0.77-1.05)	0.96 (0.85-1.10)
Respiratory	1.13 (1.02-1.25)*	1.05 (0.95-1.16)	1.13 (1.06-1.21)*	1.08 (1.02-1.14)*
Cardiovascular	1.14 (1.04-1.26)*	1.06 (0.97-1.17)	1.02 (0.96-1.10)	1.04 (0.99-1.11)
Obstetric	1.13 (0.83-1.55)	1.34 (1.06-1.70)*	1.36 (1.16-1.59)*	1.25 (1.11-1.40)*
Pediatric	1.15 (0.70-1.88)	1.32 (0.92-1.89)	1.33 (1.05-1.70)*	1.19 (1.01-1.41)*
Stroke	0.99 (0.78-1.26)	1.03 (0.81-1.31)	0.94 (0.79-1.13)	0.91 (0.79-1.06)
Assault	0.24 (0.13-0.44)*	0.28 (0.18-0.44)*	0.43 (0.33-0.57)*	0.43 (0.35-0.53)*
Mental Health	0.87 (0.79-0.95)*	0.92 (0.85-1.01)	0.93 (0.87-0.99)*	0.90 (0.85-0.95)*
Unconscious	0.46 (0.41-0.52)*	0.44 (0.38-0.49)*	0.56 (0.51-0.62)*	0.57 (0.52-0.62)*

Values are cumulative 24-hour relative risks (95% CI) from quasi-Poisson DLNM comparing 5 mm precipitation to 0 mm reference. Asterisk (*) indicates $P < 0.05$. Race composition quartiles based on ZIP code percent non-white population from ACS 2016-2020 5-year estimates; mean percent non-white shown in column headers. Unlike income-stratified precipitation effects (Extended Data Figure 2; respiratory significant only in Q1-Q2), race composition does not show a monotonic gradient: respiratory effects are significant in Q1 (whitest), Q3, and Q4 but not Q2; cardiovascular effects are significant only in Q1. Obstetric and pediatric effects peak in Q2-Q3 rather than Q4. This non-monotonic pattern suggests that precipitation-related EMS demand is more strongly structured by income than by racial composition. Stratum-hour aggregated DLNM (population-weighted weather within quartile), adjusted for temperature (ns, 4 df), seasonality (ns, 6 df), trend (ns, 4 df/year), hour, day-of-week, COVID-19 acute period, $PM_{2.5}$, O_3, NO_2. n=17.9 million ZIP-hour observations across 170 ZIP codes, New York City, 2012-2023.

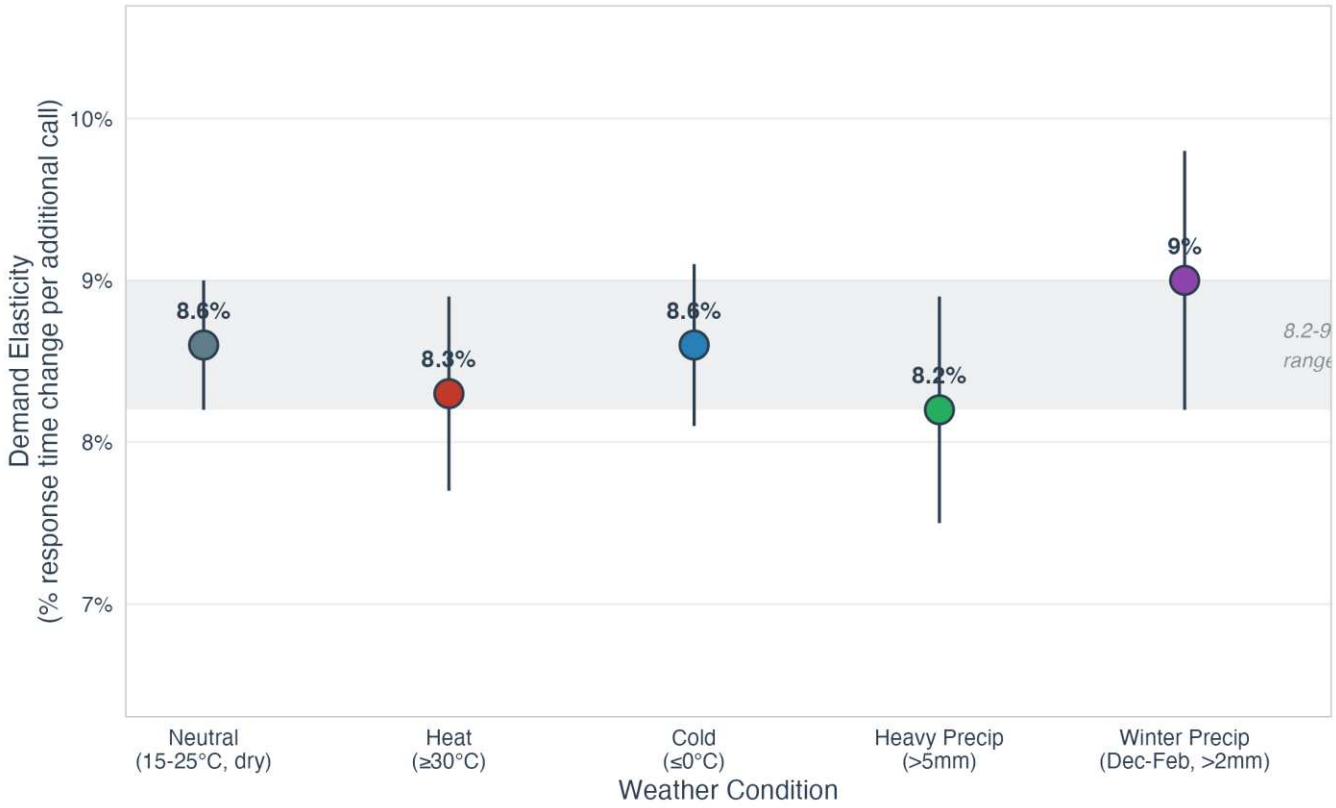

Total Calls
Cardiac Arrest
Stroke
Cardiovascular
Respiratory
Trauma
Assault
Mental Health
Poisoning/Overdose
Unconscious
Seizures
Obstetric
Gastrointestinal
Allergic
Bleeding
Hypertensive
Pediatric
Burns
Non-Specific
Relative Risk
Temperature (°C)



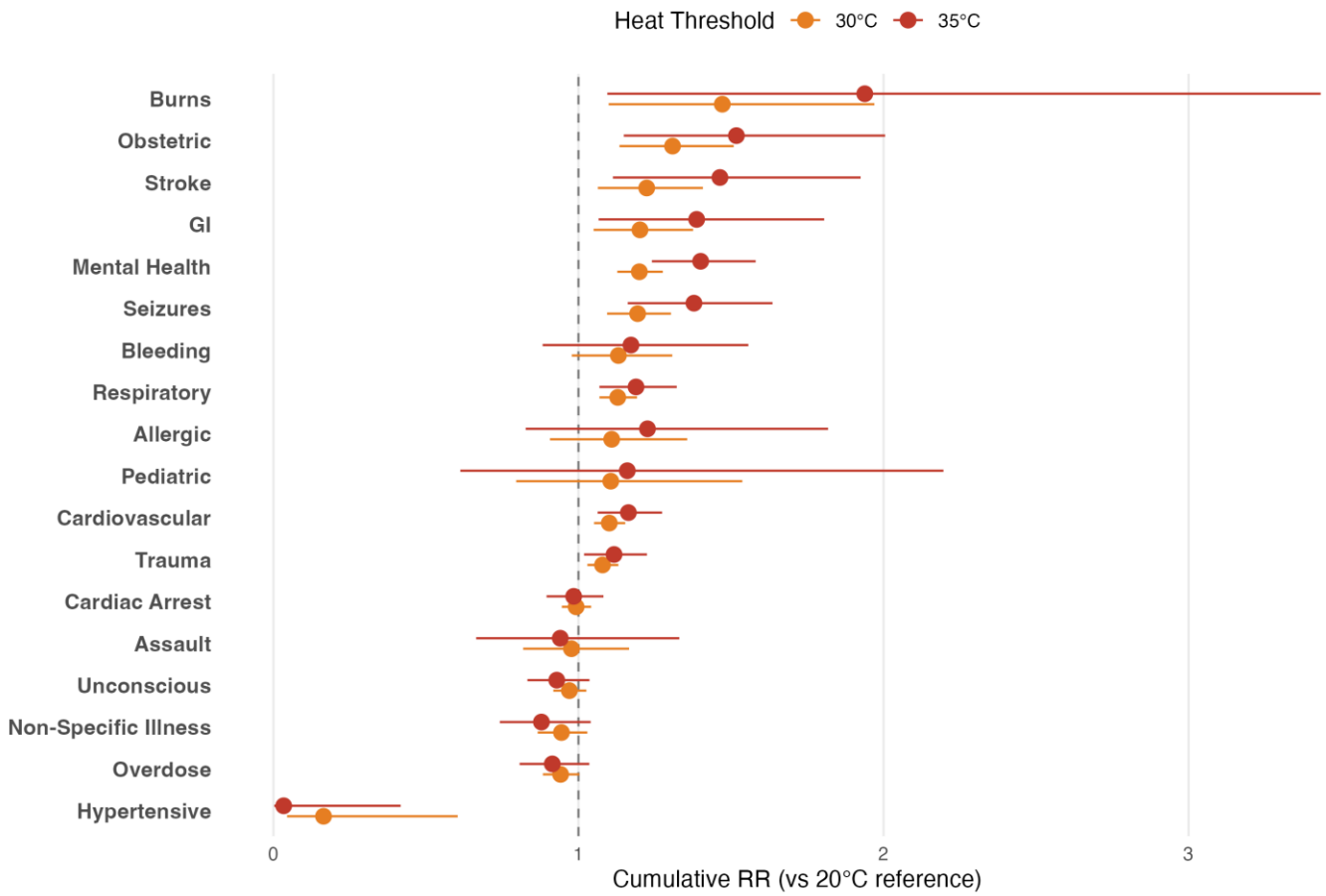
Extended Data Figure 1. Temperature and precipitation dose-response relationships across all 19 clinical endpoints. Main figures present 10 core endpoints; these panels show complete results. (A) Temperature dose-response curves showing cumulative 24-hour relative risk for each endpoint (reference: 20°C). Most endpoints show J-shaped relationships with effects increasing above 25°C. Unconscious/Syncope and Burns show the steepest increases at high temperatures; Hypertensive emergencies show a unique inverse relationship. (B) Precipitation dose-response curves (reference: 0 mm, truncated at 5 mm). Precipitation effects are generally smaller than temperature effects. Allergic reactions decreased during precipitation; Trauma increased at higher precipitation. Shaded regions indicate 95% confidence intervals. DLNM with natural cubic splines (4 df temperature, 3 df lag). n=105,192 city-hour observations, New York City, 2012-2023.



Extended Data Figure 2. Socioeconomic effect modification of weather effects on EMS call volume. Dumbbell plot showing heat (30°C vs 20°C), cold (0°C vs 20°C), and heavy rain (5–10 mm vs none) effects by income quartile (Q1=lowest, Q4=highest) for 10 core clinical endpoints. Temperature effects estimated from DLNM crossbasis; precipitation effects from categorical quasi-Poisson models with temperature adjustment. Points represent rate ratios; connecting lines illustrate Q1-Q4 disparity magnitude. Error bars show 95% confidence intervals. Heat effects are uniform across income quartiles (~10-11%; heterogeneity $P = 0.44$); precipitation shows income-graded effects concentrated in lower-income neighborhoods (Q1–Q2). Race × Temperature stratification results available in supplementary materials. n=17.8 million ZIP-hour observations (169 ZIP codes with valid income data), New York City, 2012-2023.

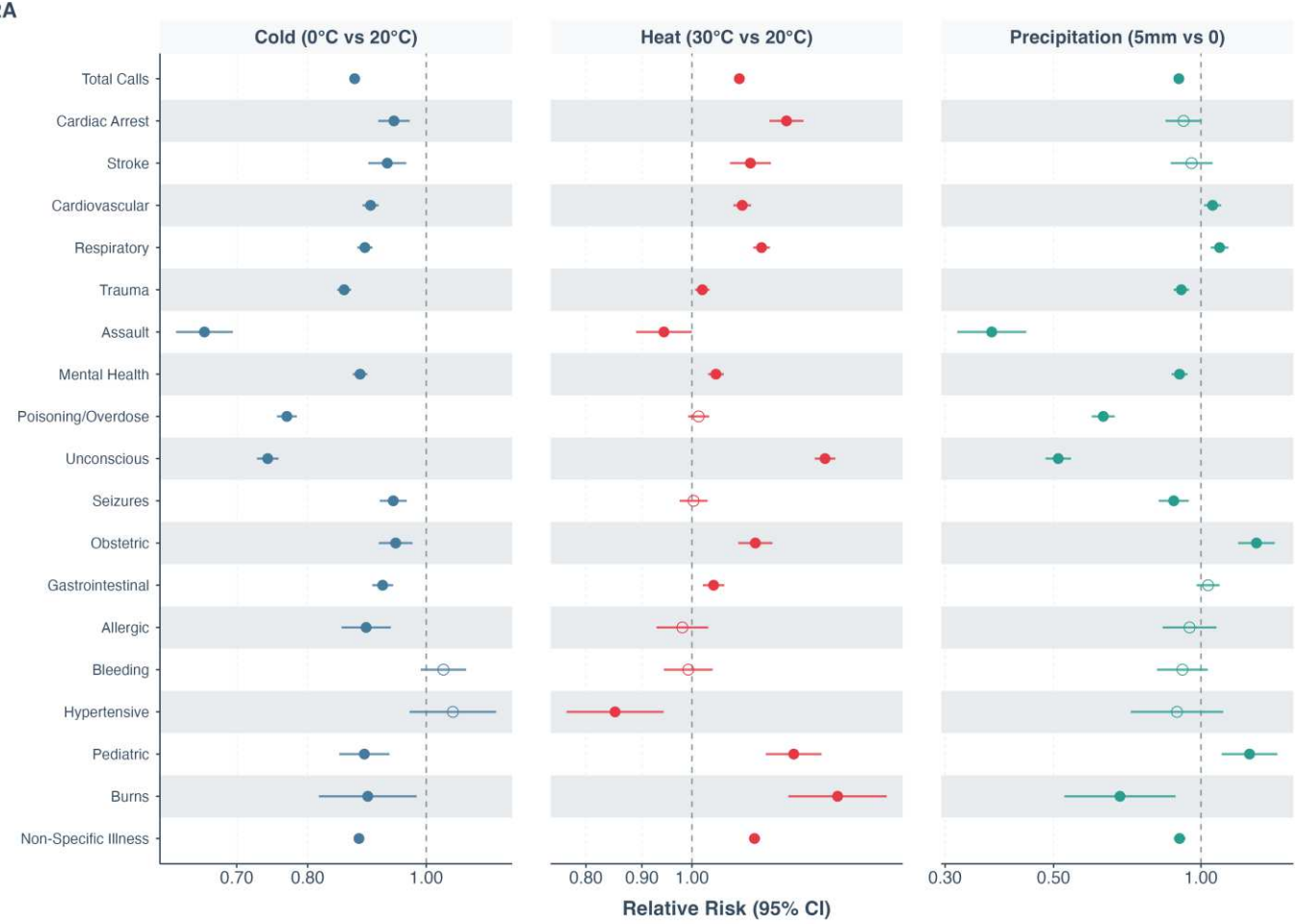


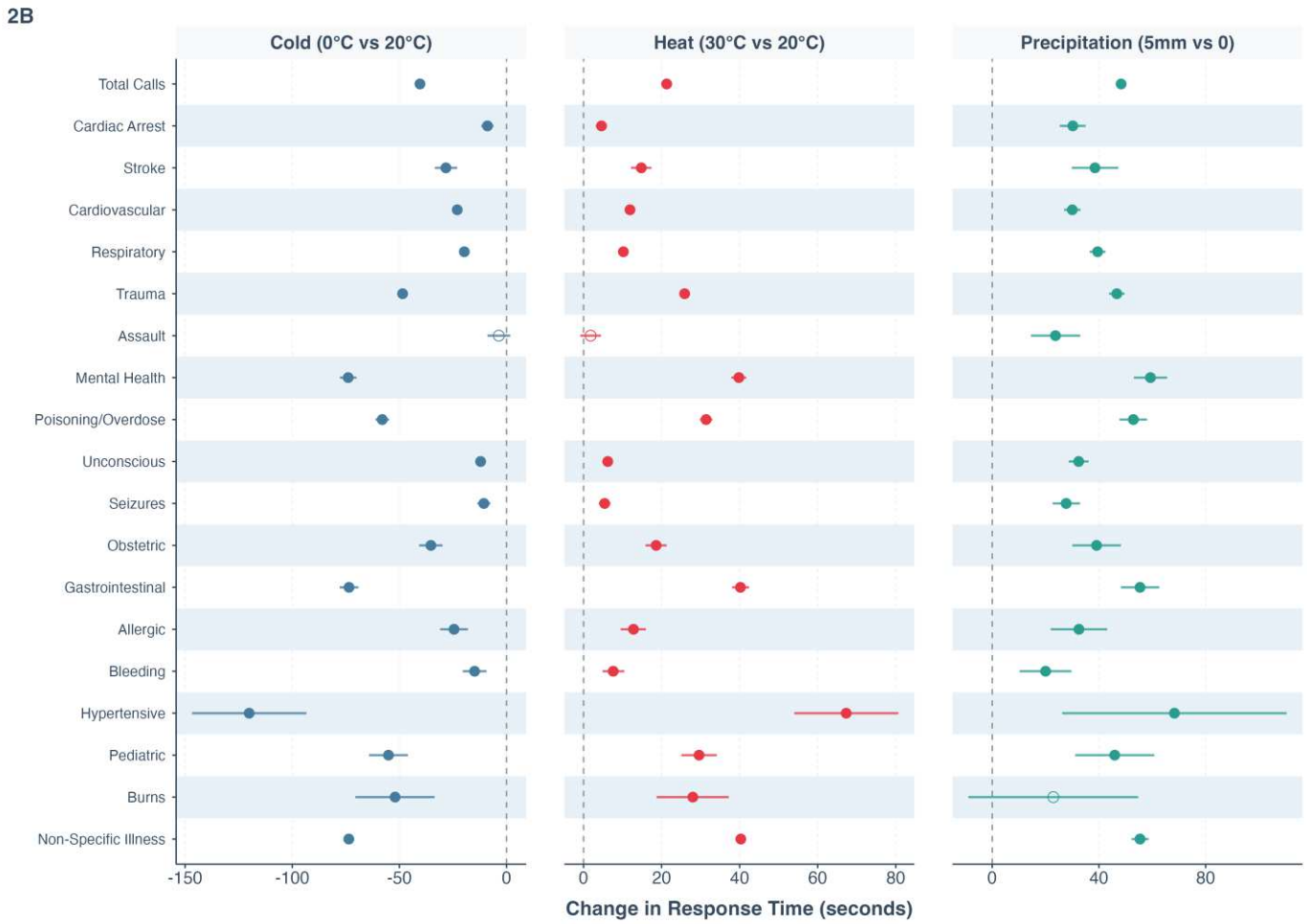
Extended Data Figure 3. Demand elasticity constancy across weather conditions. Demand elasticity (percent change in response time per additional EMS call) by weather condition: neutral (15-25°C, dry), heat (≥30°C), cold (≤0°C), heavy precipitation (>5 mm), and winter precipitation (December–February, >2 mm). The range (8.8-10.1%) across all conditions is consistent with weather effects on response time operating through direct operational degradation rather than demand-mediated queueing congestion. Gray band shows the 8.8-10.1% range. If heat-related delays were demand-driven, elasticity would increase during adverse weather; constant elasticity supports independent mechanisms. n=9.1 million ZIP-hour observations with response time data, New York City, 2012-2023.



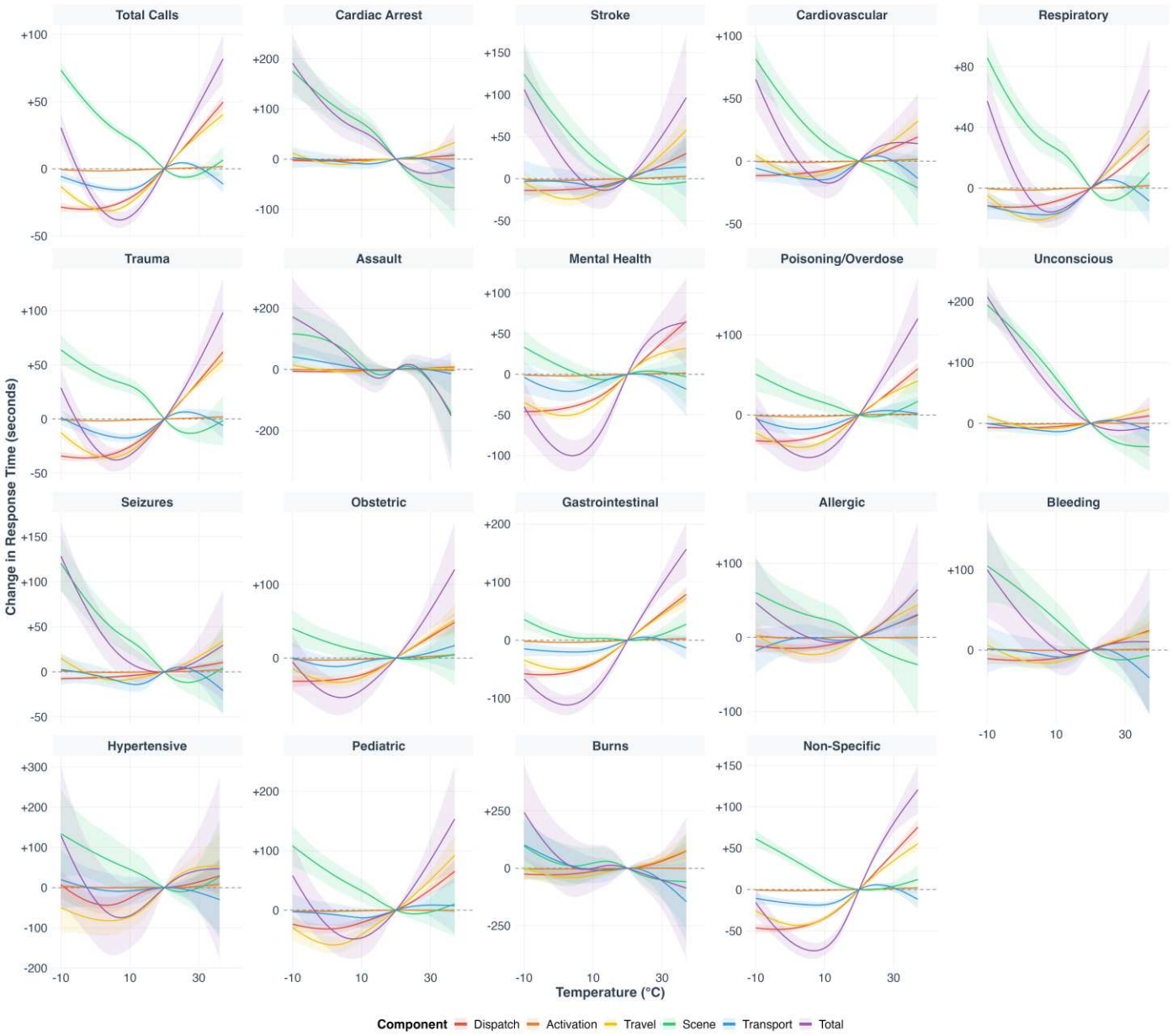
Extended Data Figure 4. Dose-response in severity under-triage by heat threshold (DLNM). Cumulative 24-hour relative risk of severity under-triage (initial severity level > final severity, indicating underestimation) at 30°C (orange) and 35°C (red) versus 20°C reference, by clinical endpoint. Effects intensify at 35°C: stroke under-triage increases from RR 1.22 to 1.46, burns from 1.47 to 1.94, obstetric from 1.31 to 1.52, mental health from 1.20 to 1.40, seizures from 1.19 to 1.38, respiratory from 1.13 to 1.19, and cardiovascular from 1.10 to 1.16. Cardiac arrest remains null at both thresholds (30°C: RR 0.99; 35°C: 0.98). Quasi-Poisson DLNM with offset(log valid severity observations/hour), adjusted for hour, day-of-week, seasonality, trend, $PM_{2.5}$, O_3, NO_2. Population-weighted city-hour weather. n=16.5 million incidents, New York City, 2012-2023.

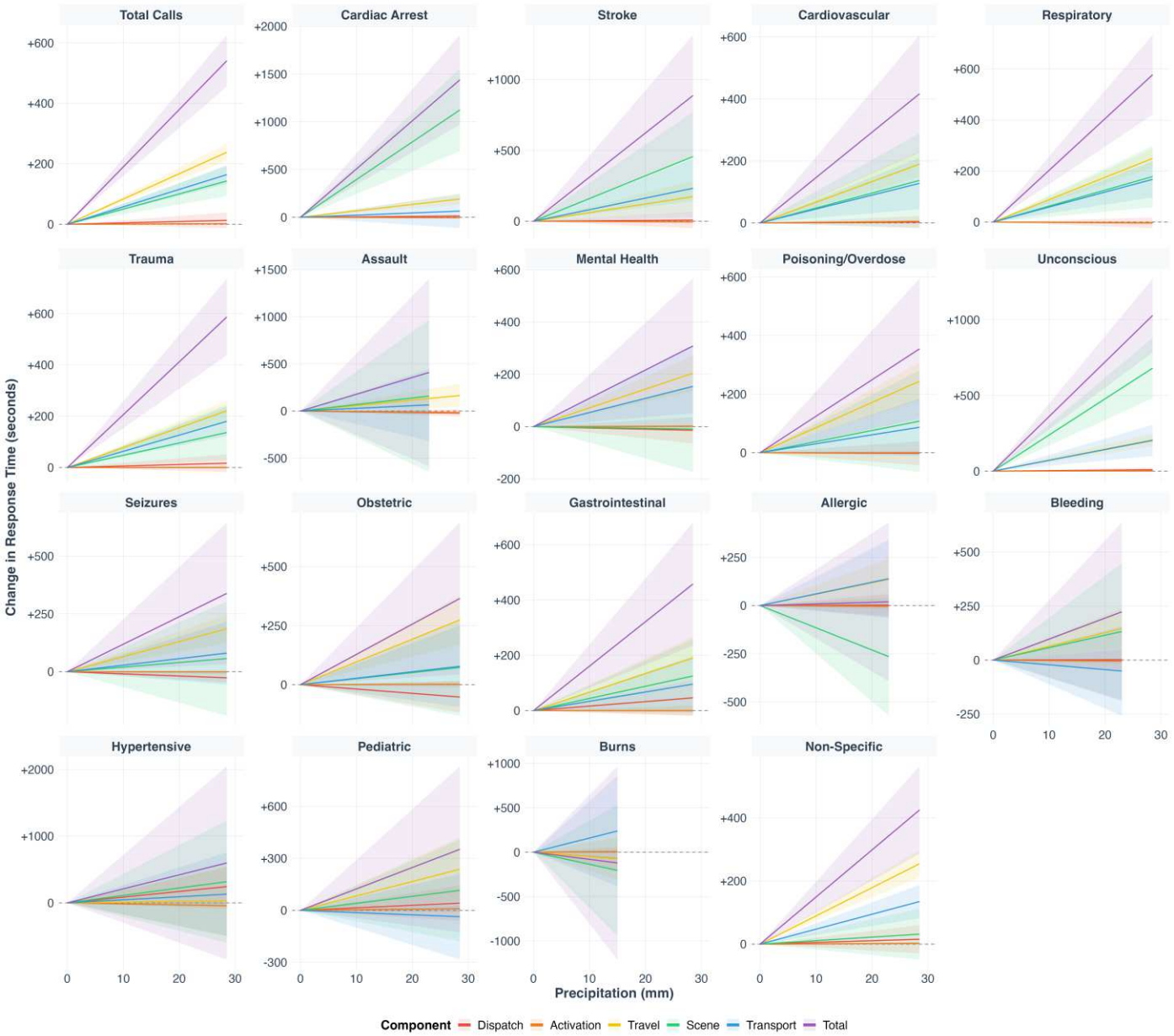
2A


Cold (0°C vs 20°C)
Heat (30°C vs 20°C)
Precipitation (5mm vs 0)
Total Calls
Cardiac Arrest
Stroke
Cardiovascular
Respiratory
Trauma
Assault
Mental Health
Poisoning/Overdose
Unconscious
Seizures
Obstetric
Gastrointestinal
Allergic
Bleeding
Hypertensive
Pediatric
Burns
Non-Specific Illness
0.70
0.80
1.00
0.80
0.90
1.00
0.30
0.50
1.00
Relative Risk (95% CI)



Extended Data Figure 5. Weather effects on EMS call volume and response time across all 19 clinical endpoints. Main figures (Figures 2, 4) present 10 core endpoints selected for narrative roles; these panels show complete results. (A) Forest plot showing cumulative 24-hour relative risks for heat (30°C vs 20°C), cold (0°C vs 20°C), and precipitation (5 mm vs 0 mm) effects on call volume. Burns showed the largest heat effect (RR 1.36), followed by Unconscious/Syncope (RR 1.32). Cold effects were most pronounced for Assault (RR 0.66) and Poisoning/Overdose (RR 0.77). (B) Forest plot showing weather effects on response time (seconds change from baseline). Heat increased response time across most endpoints (+21 seconds, +3.7% for total calls at 30°C). Filled circles indicate statistically significant effects ($P < 0.05$); open circles indicate non-significant effects. Error bars show 95% confidence intervals. Panel A: quasi-Poisson DLNM (n=105,192 city-hours). Panel B: incident-level log-linear models (n=16.5 million incidents). Both adjusted for hour, day-of-week, seasonality, and trend. New York City, 2012-2023.


Total Calls
Cardiac Arrest
Stroke
Cardiovascular
Respiratory
Trauma
Assault
Mental Health
Poisoning/Overdose
Unconscious
Seizures
Obstetric
Gastrointestinal
Allergic
Bleeding
Hypertensive
Pediatric
Burns
Non-Specific
Change in Response Time (seconds)
Temperature (°C)
Component
Dispatch
Activation
Travel
Scene
Transport
Total



Extended Data Figure 6. Weather effects on response time components by clinical endpoint. Cumulative effects from ZIP-hour DLNM showing response time changes across (A) temperature and (B) precipitation ranges for 6 response time components (total, dispatch, activation, travel, scene, transport) stratified by 19 clinical endpoints including Total Calls. Components include: total response (call receipt to patient contact), dispatch (call to unit assignment), activation (assignment to en route), travel (en route to scene), transport (scene departure to hospital), and scene (arrival to departure). Reference value: 20°C. Blue shading indicates cold temperatures (<10°C); red shading indicates heat (>25°C). Heat effects are most pronounced for allergic, obstetric, and hypertensive emergencies (conditions requiring clinical judgment) while protocol-driven conditions (cardiac arrest) show the smallest increases. n=9.13 million ZIP-hour observations with response time data, New York City, 2012-2023.

SUPPLEMENTARY INFORMATION

Note: Extended Data Tables 1–4 and Extended Data Figures 1–6 are presented above (10 Extended Data items). Supplementary Tables begin at S4 and Supplementary Figures at S5 to maintain distinct numbering from Extended Data items.

Supplementary Table S4. Correlations Between Temperature, Heat Index, and Dew Point (City-Hour)

Pair	Pearson r
Temperature vs Heat Index	0.990
Temperature vs Dew Point	0.919
Heat Index vs Dew Point	0.915

Correlations computed at the city-hour level using population-weighted weather variables. Heat index calculated from temperature and relative humidity using the NOAA formula; dew point computed via the Magnus approximation.

Supplementary Table S5. Heat Index Sensitivity Analyses

Outcome	Contrast	Effect (95% CI)
Total calls	Heat index 30°C vs 20°C	RR 1.23 (1.20-1.26)
Total calls	Heat index 35°C vs 20°C	RR 1.25 (1.20-1.29)
Total response time	Heat index 30°C vs 20°C	+8.9% (8.2-9.6%), +50.5 s
Total response time	Heat index 35°C vs 20°C	+11.3% (10.2-12.4%), +64.4 s

Heat index sensitivity models use the same DLNM specification as primary analyses, substituting heat index for temperature. The heat index scale amplifies nonlinearly at high humidity, compressing a wider range of physiological stress into the same numeric contrast; accordingly, heat index RR magnitudes are larger than dry-bulb temperature estimates at the same numeric threshold and the two scales are not directly comparable. Response time effects are percent changes from log-linear DLNM with call-volume weights. Population-weighted city-hour weather from NLDAS-2.

Supplementary Table S6. Extreme Heat and Moderate Cold Effects on EMS Call Volume by Endpoint

Endpoint	Extreme Heat (35°C vs 20°C)	Extreme Cold (-5°C vs 20°C)
Total Calls	1.16 (1.14-1.18)	0.87 (0.86-0.88)
Cardiac Arrest	1.42 (1.32-1.52)	0.99 (0.96-1.02)
Stroke	1.22 (1.12-1.33)	0.93 (0.89-0.97)
Cardiovascular	1.20 (1.15-1.24)	0.89 (0.87-0.90)
Respiratory	1.28 (1.23-1.32)	0.90 (0.88-0.91)
Trauma	0.98 (0.95-1.01)	0.85 (0.83-0.86)
Assault	0.80 (0.71-0.90)	0.62 (0.58-0.66)
Mental Health	1.06 (1.03-1.10)	0.86 (0.85-0.88)
Poisoning/Overdose	0.95 (0.91-1.00)	0.73 (0.72-0.75)
Unconscious	1.56 (1.49-1.63)	0.73 (0.71-0.74)
Seizures	1.00 (0.94-1.06)	0.91 (0.89-0.94)
Obstetric	1.25 (1.16-1.34)	0.94 (0.91-0.98)
Gastrointestinal	1.07 (1.02-1.12)	0.90 (0.88-0.92)
Allergic	0.92 (0.83-1.03)	0.88 (0.83-0.93)
Bleeding	1.01 (0.91-1.11)	1.04 (1.00-1.09)
Hypertensive	0.80 (0.65-0.98)	1.07 (0.97-1.17)
Pediatric	1.43 (1.27-1.60)	0.86 (0.81-0.91)
Burns	1.57 (1.28-1.92)	0.95 (0.86-1.05)

Values are cumulative 24-hour relative risks (95% CI) from quasi-Poisson DLNM comparing temperatures to 20°C reference. Extreme heat (35°C) occurs <0.2% of hours (~17 hours/year in NYC); moderate cold (-5°C) occurs ~3% of hours. Models adjusted for hour, day-of-week, seasonality (6 df), long-term trend (4 df/year), and air pollutants ($PM_{2.5}$, O_3, NO_2). Population-weighted city-hour weather from NLDAS-2. n=16.5 million incidents, New York City, 2012-2023.

Supplementary Methods: Queueing Diagnostics

High call volumes may create system congestion that delays response to subsequent calls, generating temporal spillover effects predicted by queueing theory (Hu et al., 2018; Green et al., 2006). We tested this using DLNM with call volume as the exposure and response time as the outcome, specifying a 12-hour lag structure (natural splines with 4 df for exposure, 3 df for lag). We report cumulative effects at the 90th percentile of call volume (219 calls/hour) compared to median (165 calls/hour), decomposed into contemporaneous (lag 0) and spillover (lags 1-12) components. To validate queueing theory predictions, we stratified analyses by capacity utilization: low (<80%), normal (80-100%), high (100-120%), and saturated (>120%). EMS systems can be conceptualized as M/M/c queues, where arrivals (911 calls) follow a Poisson process (first M), service times (incident handling duration) are exponentially distributed (second M), and c represents the number of parallel servers (available ambulance crews) (Hu et al., 2018). Under M/M/c theory, mean waiting time W increases non-linearly as system utilization ρ (arrival rate divided by service capacity) approaches 1, following $W \propto 1/(1-\rho)$. This predicts that spillover effects should increase convexly as utilization approaches saturation, a relationship we explicitly test through capacity-stratified analyses.

Supplementary Results: Temporal and Seasonal Effect Modification

Time-of-Day Stratification: Time-of-day stratified analysis revealed that apparent "delayed" effects in pooled models are artifacts of temporal mixing rather than prolonged physiological processes. Daytime incidents (7:00-18:59) showed contemporaneous effects with peak at lag 0 hours, consistent with immediate heat stress responses. Nighttime incidents (19:00-6:59) showed lagged effects with peak at 10-12 hours, reflecting afternoon heat exposure producing health events that present during evening and overnight hours. This temporal structure has important implications for heat warning systems: afternoon heat exposure creates health risks extending into nighttime hours even after ambient temperatures decline.

Precipitation showed a contrasting pattern: true 12-24 hour lags across all time periods regardless of when incidents occurred. This consistent lag structure (unlike the time-dependent pattern for heat) supports the interpretation that our method can distinguish temporal mixing from true delayed effects and suggests that precipitation effects operate through different pathways (e.g., delayed infrastructure impacts, flooding, delayed care-seeking) than heat effects.

COVID-19 Natural Experiment: COVID-19 lockdowns coincided with multiple behavioral and system changes. Heat effects were modestly amplified during the COVID period (total calls at 30°C: RR 1.19 during COVID vs 1.13 pre-COVID), returning toward baseline post-COVID (RR 1.17), with cardiac arrest showing the largest amplification (RR 1.44 vs 1.09). Concurrent shifts in care-seeking, acuity mix, and system operations preclude clean attribution, but the pattern is consistent with pandemic-related system stress rather than reduced outdoor exposure attenuating heat burden.

Seasonal Acclimatization: We employed a percentile-based approach comparing season-specific 95th percentile temperatures to seasonal medians (Table S7), ensuring comparisons use temperatures that actually occur within each season. This approach showed the following patterns: Unconscious/Syncope showed consistently large heat effects across all seasons (RR 1.06-1.43), supporting direct physiological mechanisms. Cardiac arrest effects were elevated in spring (RR 1.17 for 22°C vs 10°C) and summer (RR 1.25 for 31°C vs 24°C) but minimal in winter (RR 1.03 for 11°C vs 2°C), consistent with heat-specific rather than general temperature stress. The spring elevation at relatively moderate temperatures (22°C) compared to summer's higher absolute threshold (31°C) supports pre-acclimatization vulnerability (Balmain et al., 2018; Kenney & Munce, 2003). Cold effects (Table S8) were generally protective across endpoints, except for cardiac arrest which showed increased risk during extreme winter cold (RR 1.06 for -7°C vs 2°C), consistent with cold-induced cardiovascular strain diverging from the general protective pattern. Early-season heat events may pose substantially greater risk than equivalent temperatures later in summer, suggesting heat warning thresholds should be seasonally adjusted with lower thresholds in spring.

Supplementary Table S7. Seasonal Heat Effects on EMS Call Volume (95th Percentile vs Median Temperature)

Endpoint	Winter (11°C vs 2°C)	Spring (22°C vs 10°C)	Summer (31°C vs 24°C)	Fall (26°C vs 15°C)
Total Calls	1.04 (1.03-1.05)*	1.15 (1.13-1.16)*	1.13 (1.11-1.15)*	1.13 (1.12-1.14)*
Cardiac Arrest	1.03 (0.99-1.07)	1.17 (1.11-1.23)*	1.25 (1.16-1.34)*	1.11 (1.06-1.17)*
Stroke	1.04 (0.99-1.09)	1.06 (1.00-1.12)	1.02 (0.94-1.11)	1.12 (1.06-1.19)*
Cardiovascular	1.05 (1.03-1.07)*	1.05 (1.02-1.08)*	1.09 (1.06-1.13)*	1.10 (1.07-1.13)*
Respiratory	1.05 (1.03-1.07)*	1.11 (1.09-1.14)*	1.23 (1.19-1.27)*	1.13 (1.10-1.15)*
Trauma	1.02 (1.00-1.04)*	1.20 (1.17-1.22)*	1.07 (1.04-1.10)*	1.12 (1.10-1.15)*
Assault	1.13 (1.05-1.22)*	1.38 (1.27-1.51)*	1.13 (1.01-1.27)*	1.20 (1.10-1.32)*
Mental Health	1.06 (1.04-1.08)*	1.10 (1.08-1.13)*	1.04 (1.00-1.07)*	1.12 (1.09-1.15)*
Poisoning/Overdose	1.05 (1.03-1.08)*	1.22 (1.18-1.25)*	1.03 (0.99-1.07)	1.20 (1.16-1.23)*
Unconscious	1.06 (1.03-1.09)*	1.42 (1.37-1.46)*	1.43 (1.37-1.50)*	1.32 (1.28-1.36)*
Seizures	1.00 (0.97-1.04)	1.02 (0.97-1.06)	1.04 (0.98-1.10)	1.03 (0.99-1.07)
Obstetric	1.04 (1.00-1.08)	1.00 (0.94-1.05)	1.13 (1.05-1.21)*	1.16 (1.10-1.22)*
Gastrointestinal	1.06 (1.03-1.08)*	1.03 (1.00-1.07)	1.03 (0.99-1.08)	1.11 (1.07-1.14)*
Allergic	1.06 (0.99-1.12)	1.11 (1.03-1.20)*	0.88 (0.79-0.98)*	1.14 (1.05-1.23)*
Bleeding	0.98 (0.93-1.03)	0.97 (0.91-1.05)	1.00 (0.90-1.11)	0.98 (0.91-1.06)
Hypertensive	1.08 (0.97-1.21)	0.84 (0.73-0.97)*	0.92 (0.76-1.12)	0.82 (0.72-0.94)*
Pediatric	1.02 (0.97-1.08)	1.18 (1.09-1.28)*	1.27 (1.12-1.43)*	1.13 (1.03-1.23)*
Burns	1.03 (0.91-1.16)	1.30 (1.12-1.51)*	1.12 (0.90-1.38)	1.32 (1.13-1.54)*
Non-Specific	1.03 (1.01-1.04)*	1.14 (1.12-1.16)*	1.15 (1.12-1.17)*	1.13 (1.11-1.15)*

*Values are relative risks (95% CI) from quasi-Poisson DLNM comparing season-specific 95th percentile temperature to seasonal median. This approach uses temperatures that actually occur within each season: Winter compares 11°C (warm winter day) to 2°C (typical); Summer compares 31°C (hot day) to 24°C (typical). Asterisk (*) indicates $P < 0.05$.

Unconscious/Syncope shows the largest and most consistent heat effects across seasons (RR 1.06-1.43), supporting direct heat-stress mechanisms. Spring shows elevated effects for assault (RR 1.38) and poisoning/overdose (RR 1.22), consistent with behavioral and pre-acclimatization mechanisms. Hypertensive emergencies show paradoxical decreases with heat. Models adjusted for hour, day-of-week, day-of-year (4 df), trend, within season. n=16.5 million incidents, New York City, 2012-2023.*

Supplementary Table S8. Seasonal Cold Effects on EMS Call Volume (5th Percentile vs Median Temperature)

Endpoint	Winter (-7°C vs 2°C)	Spring (0°C vs 10°C)	Summer (16°C vs 24°C)	Fall (2°C vs 15°C)
Total Calls	0.99 (0.98-0.99)*	0.95 (0.94-0.96)*	0.96 (0.94-0.97)*	0.98 (0.96-0.99)*
Cardiac Arrest	1.06 (1.02-1.10)*	0.95 (0.91-1.00)*	0.98 (0.92-1.05)	0.97 (0.93-1.02)
Stroke	1.00 (0.95-1.05)	1.00 (0.94-1.06)	0.88 (0.81-0.95)*	0.97 (0.92-1.03)
Cardiovascular	0.97 (0.95-0.99)*	0.92 (0.89-0.94)*	0.97 (0.94-1.00)	0.97 (0.94-0.99)*
Respiratory	0.99 (0.98-1.01)	0.92 (0.90-0.94)*	0.93 (0.90-0.96)*	0.95 (0.93-0.97)*
Trauma	0.99 (0.97-1.00)	1.01 (0.98-1.03)	0.97 (0.95-1.00)*	0.97 (0.95-0.99)*
Assault	0.87 (0.80-0.94)*	0.89 (0.82-0.97)*	0.99 (0.88-1.10)	0.93 (0.85-1.01)
Mental Health	0.98 (0.96-1.00)	0.98 (0.96-1.00)	0.96 (0.93-0.99)*	0.97 (0.95-1.00)*
Poisoning/Overdose	0.89 (0.87-0.91)*	0.93 (0.90-0.95)*	0.91 (0.88-0.95)*	0.90 (0.88-0.93)*
Unconscious	0.93 (0.90-0.96)*	0.89 (0.86-0.91)*	0.93 (0.89-0.97)*	0.93 (0.90-0.96)*
Seizures	0.92 (0.89-0.95)*	0.94 (0.91-0.98)*	1.03 (0.97-1.08)	0.98 (0.94-1.02)
Obstetric	1.01 (0.97-1.05)	0.96 (0.91-1.01)	0.91 (0.85-0.98)*	0.97 (0.93-1.03)
Gastrointestinal	1.00 (0.98-1.03)	0.97 (0.94-1.00)*	0.97 (0.93-1.02)	0.99 (0.96-1.02)
Allergic	1.03 (0.96-1.09)	0.98 (0.91-1.06)	0.92 (0.83-1.01)	1.04 (0.96-1.12)
Bleeding	1.01 (0.96-1.07)	0.97 (0.91-1.04)	1.05 (0.95-1.16)	1.00 (0.94-1.07)
Hypertensive	1.03 (0.92-1.16)	1.08 (0.95-1.23)	1.21 (1.01-1.46)*	0.94 (0.83-1.08)
Pediatric	0.99 (0.93-1.04)	1.03 (0.95-1.11)	0.95 (0.85-1.07)	1.05 (0.98-1.13)
Burns	1.15 (1.02-1.30)*	1.12 (0.97-1.30)	0.72 (0.58-0.88)*	1.19 (1.03-1.38)*
Non-Specific	0.97 (0.96-0.98)*	0.92 (0.91-0.94)*	0.95 (0.93-0.98)*	1.01 (0.99-1.03)

Values are relative risks (95% CI) from quasi-Poisson DLNM comparing season-specific 5th percentile temperature to seasonal median. This approach uses temperatures that actually occur within each season: Winter compares -7°C (cold winter day) to 2°C (typical); Summer compares 16°C (cool summer day) to 24°C (typical). Asterisk (*) indicates $P < 0.05$. Cold effects are generally protective (RR <1) for most endpoints, reflecting reduced outdoor activity. Unconscious/Syncope shows consistent cold-protective effects across all seasons (RR 0.89-0.93). Cardiac arrest shows increased risk during extreme winter cold (RR 1.06), diverging from the general protective pattern, consistent with cold-induced cardiovascular strain. Poisoning/Overdose and Assault show strong cold-protective effects, consistent with behavioral mechanisms. Burns show paradoxical patterns: increased in winter cold (indoor heating fires) but decreased in summer cold. Models adjusted for hour, day-of-week, day-of-year (4 df), trend, within season. n=16.5 million incidents, New York City, 2012-2023.

Supplementary Table S9. COVID-19 Period Heat Effects on EMS Call Volume by Endpoint (30°C vs 20°C)

Endpoint	Pre-COVID (2012-2019)	COVID (2020-2021)	Post-COVID (2022-2023)
Total Calls	1.13 (1.12-1.13)*	1.19 (1.18-1.21)*	1.17 (1.16-1.19)*
Cardiac Arrest	1.09 (1.05-1.13)*	1.44 (1.34-1.55)*	1.16 (1.07-1.25)*
Respiratory	1.15 (1.14-1.17)*	1.39 (1.33-1.45)*	1.05 (1.01-1.10)*
Mental Health	0.99 (0.97-1.01)	1.10 (1.07-1.14)*	1.07 (1.04-1.11)*
Assault	1.25 (1.19-1.31)*	1.23 (1.09-1.39)*	1.26 (1.13-1.42)*
Overdose	1.09 (1.07-1.12)*	1.14 (1.08-1.19)*	1.21 (1.15-1.27)*

Values are relative risks (95% CI) from quasi-Poisson DLNM comparing 30°C to 20°C reference. COVID period (2020-2021) included pandemic-related behavioral changes and service disruptions. Elevated cardiac arrest effects during COVID may reflect delayed care-seeking and pandemic stress. Mental health heat effects were null pre-COVID (RR 0.99) but significant during COVID (RR 1.10) and post-COVID (RR 1.07), likely reflecting pandemic-era increases in mental health call volume and shifting case mix rather than altered heat sensitivity. Asterisk (*) indicates $P < 0.05$. Models adjusted for hour, day-of-week, seasonality, long-term trend, and air pollutants. n=16.5 million incidents, New York City, 2012-2023.

Supplementary Table S10. Response Time Timestamp Missingness by Temperature Bin

Temperature Bin (°C)	N Incidents	Missing (%)
< -5	313,156	3.57
-5 to <0	880,081	3.50
0–5	2,113,610	3.38
5–10	2,712,390	3.50
10–15	2,529,556	3.50
15–20	2,458,306	3.72
20–25	3,081,715	3.90
25–30	2,185,240	4.16
30–35	549,351	4.55
≥35	15,160	5.59

Note: Missingness = 1 − valid response time. N reflects all incidents with weather-matched ZIP codes (170 ZIP codes) regardless of income data availability; the total differs from the 16,486,293 in Table 1, which excludes 1 ZIP code lacking ACS income data. Aggregated-bin logistic regression of valid response time (outcome=1 if present) yields OR per 10°C = 0.93 (95% CI 0.92-0.93), indicating slightly higher missingness at higher temperatures, though absolute differences are small (3.4% at 0–5°C to 5.6% at ≥35°C).

Supplementary Table S11. Response Time Timestamp Missingness by Precipitation Bin

Precipitation Bin (mm/hour)	N Incidents	Missing (%)
>0–0.1	15,407,386	3.70
0.1–2.5	1,232,401	3.82
2.5–5	140,382	4.01
5–10	50,818	4.02
≥10	7,578	4.02

Note: Aggregated-bin logistic regression of valid response time (outcome=1 if present) yields OR per 5 mm = 0.91 (95% CI 0.90-0.93), indicating slightly higher missingness with higher precipitation, though absolute differences are small (3.7% at ≤0.1 mm to 4.0% at ≥5 mm).

Supplementary Table S12. Sensitivity: Non-specific Classification Excluding Syncope

Endpoint	RR (30°C vs 20°C)	RR (35°C vs 20°C)
Stroke	1.11 (0.98-1.27)	1.14 (0.87-1.48)
Cardiac Arrest	1.11 (0.97-1.27)	1.16 (0.89-1.51)
Mental Health	1.13 (1.06-1.21)	1.25 (1.10-1.42)
Respiratory	1.10 (1.02-1.19)	1.11 (0.96-1.29)
Obstetric	1.13 (0.93-1.37)	1.29 (0.88-1.89)

Note: Non-specific pattern excludes SYNCOP codes; models otherwise match primary DLNM specification and use incident-weighted city-hour weather.

Supplementary Table S13. Long-term Temporal Trends in Call Volume and Response Time

Period	Total Calls	Mean Calls / ZIP-hour	Response Time (min)	Dispatch (min)	Activation (min)	Travel (min)	Scene (min)	Transport (min)
2012–2016	6,449,790	0.865	9.16	1.33	0.36	7.43	20.92	11.17
2017–2021	7,075,274	0.950	9.27	1.00	0.44	7.82	21.39	12.33
2022–2023	2,989,111	1.004	10.82	1.63	0.49	8.56	22.95	12.58

Note: Means are call-weighted across ZIP-hours. Response time is call receipt → on-scene arrival and equals dispatch (call receipt → unit assignment) + activation delay (assignment → enroute) + travel (enroute → on-scene). Periods are non-overlapping (2012–2016, 2017–2021, 2022–2023). The total across periods (16,514,175) reflects 170 ZIP codes and differs from Table 1 (16,486,293 across 169 ZIP codes with valid ACS income data) and Supplementary Table S10 (16,838,565 across all incidents with weather-matched ZIP codes regardless of timestamp availability).

Supplementary Table S14. Residual Autocorrelation Diagnostics for Primary and Representative Stratified Models

Model	Analysis level	ACF(1)	ACF(24)	Ljung-Box *P* (24 lags)
Total calls	City-hour	0.608	0.273	<0.001
Cardiac arrest calls	City-hour	0.031	0.030	<0.001
Respiratory calls	City-hour	0.168	0.116	<0.001
Total response time (log mean, weighted)	City-hour	0.773	0.348	<0.001
Dispatch time (log mean, weighted)	City-hour	0.735	0.271	<0.001
Total calls	Income-stratum-hour (Q1)	0.479	0.228	<0.001
Total response time (log mean, weighted)	Income-stratum-hour (Q1)	0.745	0.328	<0.001

Values are from deviance-residual autocorrelation diagnostics (dlnm_diagnostics) applied to fitted models. Because the long hourly series yields high power for Ljung-Box tests, we report effect-size metrics (ACF) and interpret statistical significance together with multi-endpoint consistency and sensitivity analyses.

Supplementary Table S15. Fixed-Contrast Seasonal Heat Effects on EMS Call Volume (30°C vs 20°C)

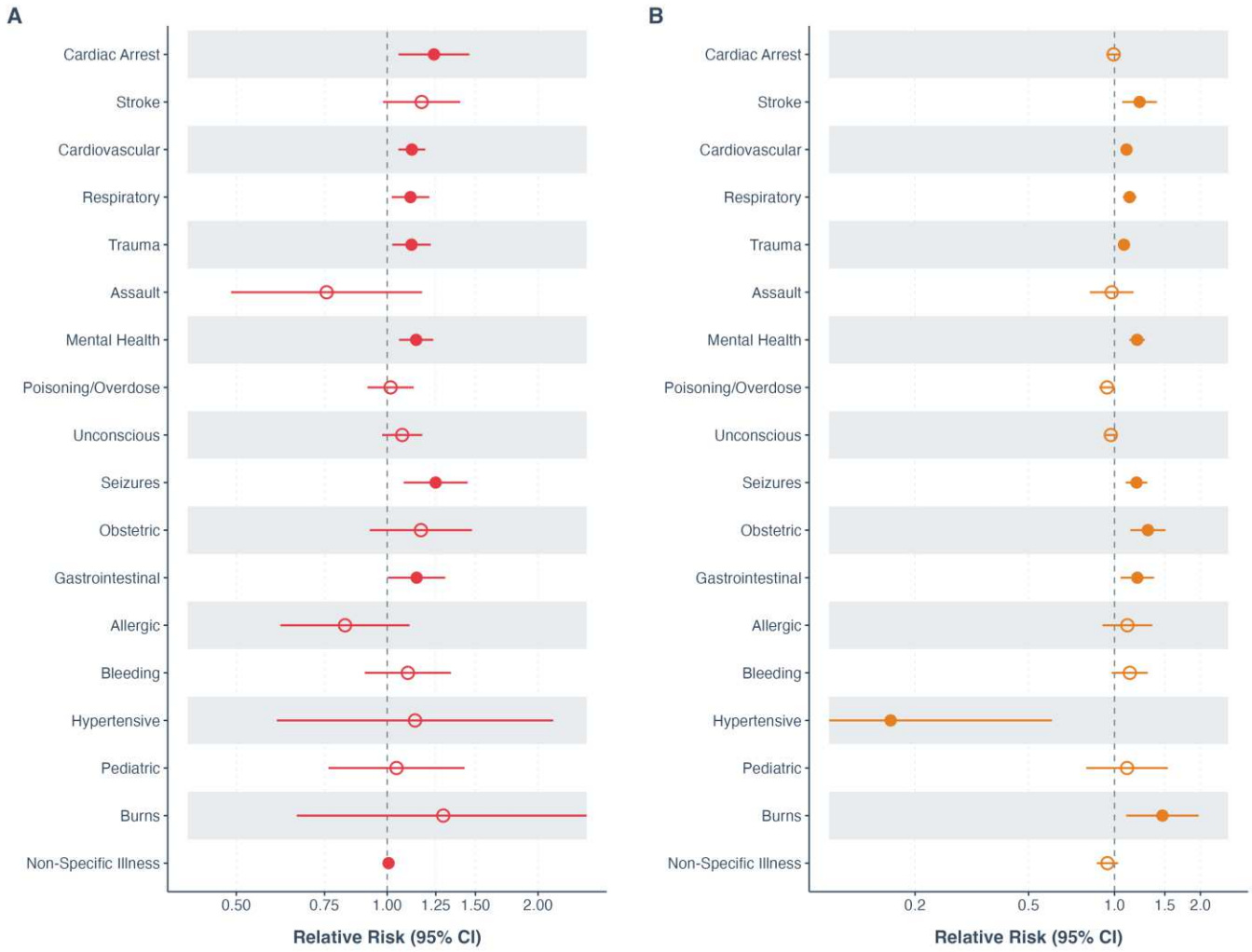
Endpoint	Winter	Spring	Summer	Fall
Total Calls	1.17 (1.12-1.22)*	1.15 (1.13-1.17)*	1.13 (1.12-1.14)*	1.11 (1.09-1.13)*
Cardiac Arrest	1.23 (0.91-1.66)	1.86 (1.62-2.14)*	1.06 (1.00-1.13)	1.11 (0.99-1.24)
Respiratory	2.44 (2.16-2.76)*	1.90 (1.79-2.01)*	1.23 (1.19-1.26)*	1.45 (1.39-1.52)*
Assault	2.31 (1.44-3.70)*	1.21 (0.98-1.50)	1.30 (1.20-1.40)*	1.25 (1.06-1.47)*
Mental Health	1.19 (1.02-1.38)*	0.87 (0.81-0.93)*	1.05 (1.02-1.08)*	0.87 (0.82-0.91)*
Overdose	2.08 (1.70-2.54)*	0.98 (0.89-1.07)	1.13 (1.09-1.17)*	1.23 (1.15-1.32)*

Values are cumulative 24-hour relative risks (95% CI) from quasi-Poisson DLNM comparing 30°C to 20°C reference within each season. This fixed-contrast approach uses the same temperature comparison across seasons; note that 30°C occurs rarely in winter and spring (primarily late May), so winter and spring estimates extrapolate from the spline fit and should be interpreted with caution. Asterisk (*) indicates $P < 0.05$. The monotonic decline in total calls RR from winter (1.17) to fall (1.11) is consistent with population-level acclimatization. The large spring cardiac arrest estimate (RR 1.86) may partly reflect spline extrapolation into sparse data. Complementary within-season percentile-based contrasts are presented in Supplementary Tables S7-S8. Models adjusted for hour, day-of-week, day-of-year, and trend within season. n=16.5 million incidents, New York City, 2012-2023.

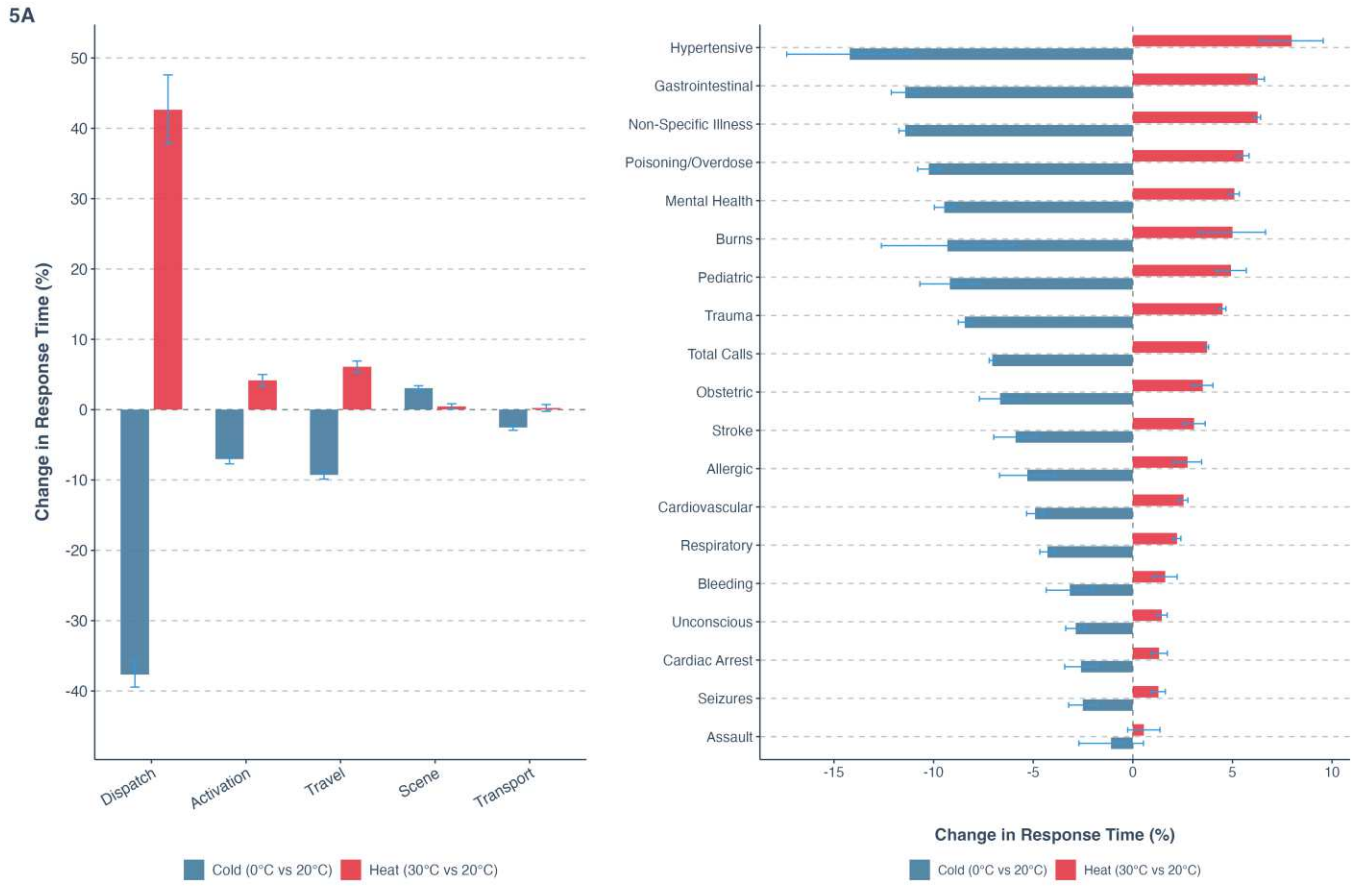
Supplementary Table S16. Seasonal Precipitation Effects on EMS Response Time (5 mm vs 0 mm)

Season	Total Response	Dispatch	Activation	Travel	Scene	Transport
Winter	+88.0%	+185.8%	+80.4%	+60.4%	+40.9%	+13.7%
Spring	+26.3%	+40.9%	+40.3%	+19.5%	+22.5%	+10.1%
Summer	+27.2%	+67.6%	+43.9%	+19.3%	+23.3%	+6.9%
Fall	+23.8%	+44.2%	+26.8%	+17.8%	+16.7%	+6.4%

Values are percent change in response time component at 5 mm precipitation vs 0 mm reference, from ZIP-hour DLNM within each season. Winter precipitation effects (+88.0% total response) are approximately 3–4× larger than non-winter seasons (approximately +24–27%), coinciding with snow, ice, and reduced road-surface quality rather than demand changes. Winter dispatch effects (+185.8%) are disproportionately large, suggesting that winter precipitation compounds information-processing delays beyond travel impairment. Models use log-linear Gaussian DLNM with call-volume weights, adjusted for temperature (ns, 4 df), seasonality (ns, 6 df), trend (ns, 4 df/year), hour, day-of-week, COVID-19 acute period, $PM_{2.5}$, O_3, NO_2. n=9.1 million ZIP-hour observations with response time data, New York City, 2012-2023.



Supplementary Figure S5. Heat effects on EMS classification patterns across all clinical endpoints. Forest plots showing relative risk (30°C vs 20°C reference) for (A) non-specific classification and (B) severity under-triage across all 18 endpoints with misclassification data (Total Calls excluded as it has no misclassification outcome). Filled circles indicate statistically significant effects (95% CI excludes 1.0); open circles indicate non-significant effects. Seizures showed the largest non-specific classification increase (RR 1.25), followed by Cardiac Arrest (1.24) and Mental Health (1.14). For severity under-triage, Burns (RR 1.47), Obstetric (1.31), and Stroke (1.22) showed the largest increases. Hypertensive emergencies showed a protective pattern for undertriage (RR 0.16), possibly reflecting heightened clinical vigilance for hypertensive crises during heat events. Cumulative 24-hour relative risk from quasi-Poisson DLNM adjusted for hour, day-of-week, seasonality, and trend. City-hour aggregated models (n=105,192 city-hours); counts derived from 16.5 million incidents. New York City, 2012-2023.



Supplementary Figure S6. Response time decomposition across all 19 clinical endpoints. (A) Percent change in response time components comparing heat (30°C) and cold (0°C) to 20°C reference (same as main Figure 4A). (B) Response time changes by clinical outcome for all 19 endpoints, ordered by heat effect magnitude. Hypertensive emergencies showed the largest heat-related response time increase, followed by Gastrointestinal and Non-Specific calls. Cold was associated with response time reductions across most endpoints, with the largest decreases for Hypertensive and Burns. Incident-level log-linear models adjusted for hour, day-of-week, seasonality, and trend; concurrent linear temperature and precipitation terms. n=16.5 million incidents, New York City, 2012-2023.

Supplementary Table S17. Analytic Units and Estimands

Analytic Unit	N	Estimand	Primary Use	Model
City-hour	105,192	System-level cumulative burden (0-24h lag)	Citywide call volume and response time	Quasi-Poisson / Gaussian DLNM
Stratum-hour	~105,000 per quartile	Within-income-stratum cumulative burden	Income-stratified call volume	Quasi-Poisson DLNM
ZIP-hour	9.1M (with RT data)	Spatial heterogeneity and demand-capacity relationship	Income-stratified response time, demand elasticity	Gaussian DLNM, linear elasticity
Incident-level	16.5M	Contemporaneous per-call association	Endpoint-specific response time, classification	Log-linear, quasi-Poisson

Note: These represent nested aggregations of the same underlying dataset. City-hour and stratum-hour models use DLNM crossbasis terms to capture non-linear and lagged effects over 0-24 hours. Incident-level models use concurrent weather exposure without lag accumulation. Magnitudes from different analytic units are not directly comparable: city-hour DLNM cumulative estimates capture system-level burden including lagged pathways (prior-hour congestion, crew fatigue), whereas incident-level concurrent estimates isolate the direct per-call weather association. The approximately 3-fold difference between cumulative (+65 seconds) and concurrent (+21 seconds) heat effects on response time quantifies the contribution of lagged pathways. Sensitivity analyses using citywide-mean exposure yielded results within 8% of ZIP-level estimates, consistent with minimal within-city temperature variation (<1°C cross-ZIP hourly standard deviation).