

Strengthening FDNY EMS Operations to Improve Clinical Outcomes

A Mayor-Directed EMS Improvement and Modernization Initiative

Submitted to:

Mayor Zohran Mamdani

Director, Mayor's Office of Management and Budget

Lillian Bonsignore, Commissioner, Fire Department of the City of New York

Speaker Julie Menin, New York City Council

Submitted by:

James T. Kerr

Former Executive Director, New York City Emergency Medical Service

Executive Summary

The problem. FDNY EMS confronts a systemwide operational-capacity challenge, not simply a response-time problem. FDNY ambulances handled more than 1.6 million medical incidents in FY2025, including more than 620,000 life-threatening emergencies. During the first four months of FY2026, average daily ambulance in-service hours fell from 8,376 to 7,986. The FY2026 Mayor's Management Report subsequently reported that average end-to-end ambulance response to life-threatening emergencies increased from 11 minutes 21 seconds in FY2025 to 13 minutes 9 seconds in FY2026—an increase of 1 minute 48 seconds. These figures refer to the interval from the initial 911 report to ambulance arrival, not merely ambulance travel time. See Source Notes 1 and 7.

Independent evidence (October 16, 2026). The NYC Comptroller's A System Under Strain reports that average ambulance response to life-threatening emergencies increased by 3 minutes 50 seconds (41.2%) between FY2015 and FY2026, while non-life-threatening response increased by 14 minutes 49 seconds and life-threatening incident demand grew 22.9%. These trends support a systemwide capacity assessment; they do not, by themselves, establish the causes of delay. [9]

Why it is happening. Some conditions are already documented: sustained high EMS workload, declining staffed ambulance-hours, paramedic vacancies, the current internal-only paramedic promotional pathway, recurring overtime pressure, the formal EMS-to-Firefighter promotional pathway, hospital turnaround burden, and worsening response performance.[1][3][4] Other plausible contributors—including the relative effects of compensation, morale, facility conditions, absenteeism, fleet downtime, deployment practices and avoidable ambulance demand—must be quantified and documented. These issues cross FDNY, OMB, OLR, DCAS, NYC Health + Hospitals and other parts of City government and cannot be solved coherently by a single entity acting alone.

The objective. Strengthen FDNY EMS operations so that New Yorkers receive the right clinical response at the right time, while creating a reliable and sustainable system for staffing, ambulance availability, fleet support, hospital interface, demand management and public accountability. FDNY and OMB should use verified CAD demand, unit-utilization, geographic coverage, workforce, hospital-turnaround, fleet-reliability and response-performance data to determine how much additional staffed

ambulance capacity is required. The ultimate measure of success should be improved clinical outcomes and operational reliability—not a fleet target or a single average response-time statistic.

The Mayor should direct an interagency EMS Improvement and Modernization Task Force to address the following priorities as a coordinated citywide program. Many of these actions should proceed simultaneously rather than as a rigid sequence.

Reduce avoidable ambulance demand without discouraging 911 use: expand clinically supervised secondary triage, telehealth, behavioral-health alternatives, frequent-user intervention and other safe pathways that maintain ambulance availability for patients who require ambulance-level care.

Recover capacity now being lost: reduce hospital turnaround, improve fleet uptime, and use predictive/dynamic deployment to return more existing ambulance-hours to street availability.

Stabilize the workforce and address morale: address the workplace conditions that influence whether EMTs and paramedics remain in EMS, including competitive and equitable compensation, vacancies, recruitment, internal career pathways, recurring overtime, condition of facilities and professional development.

Determine and fund the sustainable ambulance capacity required: use incident-level CAD demand, geographic coverage, unit availability and utilization, workforce capacity, hospital turnaround, fleet reliability and response-performance standards to calculate how many additional staffed ambulance-hours are required to meet defined reliability and clinical-response standards citywide and by borough, tour and demand period. The analysis should then determine the number, type, tour and geographic distribution of additional BLS and ALS ambulances needed. The number of ambulances will then be an output of the analysis which should also identify where the additional staffed hours are required and the reserve vehicle and workforce capacity necessary to keep those tours operating despite vacancies, leave, maintenance and hospital delay.

Consider a targeted Ultra-Time-Critical response option only if the broader system reforms and capacity expansion still leave a clinically important gap for the most time-sensitive emergencies. For the purpose of this proposal, an Ultra-Time-Critical Medical Emergency means a narrowly defined condition for which exceptionally rapid intervention can materially affect survival or serious morbidity. FDNY Medical Affairs should define the eligible cohort. A Critical Response Team/Critical Response Vehicle (CRT/CRV) would be a specialized, non-transporting clinical resource used only for that cohort and always paired with a simultaneously dispatched, fully staffed ambulance.

If performance data demonstrate that the targeted option is warranted, define and test it only after adequate base ambulance capacity is funded and operating. The initial CRT/CRV pilot should cover a deliberately limited geographic area selected through CAD analysis of eligible incident density, time-of-day patterns, travel times and partner-ambulance availability. Its size and any later expansion of the coverage area should be driven by measured performance, not by a predetermined number of vehicles.

Measure and publicly report clinical outcomes and operational reliability: publish response-time percentiles, pending-call backlogs, scheduled-versus-operated tours, in-service hours, workforce vacancies and overtime, hospital turnaround, fleet downtime, demand-management results, CRT-to-ambulance partnering reliability and appropriate patient outcomes.

Governance. Because these improvements span labor relations, budget, hiring, hospital operations, fleet, facilities, technology and clinical policy, the Mayor should sponsor an EMS Improvement and Modernization Task Force including FDNY, OMB, OLR, DCAS, NYC Health + Hospitals, the Greater New

York Hospital Association, DOHMH, EMS labor and other appropriate partners. It should be an implementation body with accountable leadership, milestones and Council oversight.

Implementation timetable. Within 30 days, convene the Task Force, establish data-sharing arrangements and agree on measures. Within 60 days, complete an initial system assessment and identify immediate actions. Within 90 days, deliver a coordinated implementation plan to the Mayor and Council identifying targets, resources and accountable organizations. Within six months, begin priority operational improvements and evidence-supported pilots. Within 12 months, report comprehensive results, including percentages meeting clinically appropriate standards by geography. Publish progress regularly thereafter.

Conditional pilot performance standards. If FDNY Medical Affairs concludes that an Ultra-Time-Critical pilot is warranted, the following should be treated as proposed operational standards for validation and testing, not as established clinical breakpoints: a clinically capable first responder at 90% of eligible incidents within six minutes; a fully staffed ambulance within eight minutes; and, when the CRT/CRV arrives first, the accompanying ambulance within three additional minutes in 90% of cases.

The six-minute metric would run from the initial 911 report to arrival of the first FDNY responder trained and equipped to provide the immediate lifesaving intervention required by that call. Depending on the clinical need, that responder could be a CFR-D fire unit, CRT/CRV, BLS ambulance or ALS ambulance. The three-minute CRT/CRV-to-ambulance interval would be an operational safeguard designed to prevent an early-treatment resource from being left on scene waiting for unavailable ambulance capability.

Central safeguard: No CRV without a fully staffed ambulance behind it.

The Proposal

The Executive Summary identifies a citywide EMS capacity problem that crosses operations, workforce policy, labor relations, hospital interface, fleet management, capital planning and public accountability. The detailed proposal below first distinguishes what the available evidence already establishes from factors that still require measurement. It then sets out coordinated actions to reduce avoidable demand, recover lost capacity, stabilize the workforce, determine the ambulance capacity the system requires, and improve transparency. A targeted Ultra-Time-Critical response is considered only as a later option if those broader reforms do not fully address the needs of the most time-sensitive patients.

System Diagnosis: A System Capacity Problem, Not Just a Response-Time Problem

The documented increase in end-to-end ambulance response time and decline in ambulance in-service hours are symptoms of a wider capacity challenge. September 2026 Local Law 119 data also show material borough variation. The diagnosis must distinguish established performance findings from factors whose contribution remains to be measured. The analysis should encompass 911 call receipt and processing, transfer to EMS, dispatch, unit assignment, turnout and travel, on-scene care, hospital transfer and return to service.

Latest geographic findings (September 2026). Local Law 119 reporting showed average end-to-end ambulance response to life-threatening emergencies of 12:44 in Brooklyn, 16:57 in the Bronx, 15:33 in

Manhattan, 15:47 in Queens and 13:26 in Staten Island. For the City's separately defined ALS measure, approximately 26% of measured incidents received an ALS ambulance in under ten minutes citywide. This measure is not all life-threatening emergencies, and ten minutes should not be assumed to be the appropriate clinical standard for every condition. See Source Note 7.

Performance clocks and clinical standards. Report the entire interval from the initial 911 call through ambulance arrival, identify each component separately, and establish clinically appropriate standards by emergency category and responding resource. Measure both the percentage meeting each standard and the 50th, 75th and 95th percentiles, including non-life-threatening calls. Separately report first clinically capable responder arrival, when relevant, rather than treating it as interchangeable with ambulance arrival.

Reconcile official response series. For FY2026, the Mayor's Management Report and the Comptroller's NYC Open Data calculation report 13:09 for end-to-end life-threatening ambulance response, whereas the Local Law 119 series reports 13:16. FDNY should publish the response-clock definitions, eligible incidents, denominators and reconciliation methodology before comparing the series. [9]

Documented Conditions

Workload and staffing mismatch: EMS demand has grown substantially over time while staffing has not kept pace. Staffing and ambulance-hours should be tied to workload and response reliability rather than historical headcount.

Falling staffed ambulance-hours: Treat operated ambulance-hours and scheduled-versus-operated tours as core management measures, by borough and tour.

Paramedic vacancies and narrow pipeline: FDNY currently describes Paramedic as a promotional opportunity available only to FDNY EMS members.[4] The internal pathway develops experienced personnel and should be preserved. The City should establish vacancy targets and evaluate a supplemental direct/lateral entry pathway for already-certified experienced New York State paramedics when the internal pipeline cannot fill documented vacancies, without displacing existing promotional opportunities.

Internal promotion to Firefighter: FDNY provides eligible EMTs a formal promotional route to Firefighter. In February 2026, 126 members of one 181-member probationary firefighter class had previously served as EMS officers, paramedics or EMTs.[3] Career mobility should continue. While one graduating class should not be treated as indicative of a long-term attrition rate, FDNY and DCAS should analyze at least five years of EMS-to-Fire promotions, forecast that movement as planned attrition, recruit and train replacements in advance, measure replacement and overtime costs and loss of experience, and strengthen EMS compensation and clinical career opportunities so remaining in EMS is a viable and sustainable long-term career choice.

Factors Requiring Quantification

Compensation, overtime and morale: compensation inequity, recurring overtime and workplace conditions are credible recruitment and retention concerns. The Task Force should quantify their relationship to vacancies, absenteeism, turnover and lost ambulance tours.

Facilities and workplace infrastructure: conduct an EMS facility-condition assessment to determine the extent to which station condition, crew amenities, staging, restocking and training space affect retention, readiness or deployment efficiency, and identify the improvements required to support the workforce and fleet.

Hospital turnaround: Expand successful hospital liaison practices and hospital-specific accountability; treat recovered ambulance-hours as additive capacity.

Fleet availability and maintenance: quantify vehicle downtime, maintenance delays, spare availability and mechanic capacity to determine the extent to which fleet constraints contribute to lost ambulance-hours; then fund the spare vehicles, preventive maintenance, parts, mechanics, replacement cycles and maintenance bays the data show are required.

Fleet-service targets versus clinical coverage. The Comptroller reports 360 ambulances in service on January 8, 2026 against a then-reported fleet-service target of 488, and 416 on September 14 after that target was reduced to 332. The City should disclose the rationale for revising the target and reconcile vehicle-readiness figures with fully staffed, deployed and immediately dispatchable 911 units. These counts do not independently measure available clinical coverage. [9]

Avoidable ambulance-intensive demand: quantify which call types and patient groups can safely be served through alternatives before setting diversion targets. Any expansion of secondary triage, telehealth, behavioral-health or frequent-user pathways should be clinically supervised and outcome-tested.

Deployment: Use current availability, predicted near-term demand, pending-call queues, travel time and hospital conditions to position resources dynamically.

If demand rises, hospital turnaround improves, telehealth reduces low-acuity transports, workforce retention changes, or predictive deployment materially improves unit availability, the required ambulance number should change accordingly. The capacity model should be rerun at least annually and whenever there is a material change in demand, staffing, hospital turnaround or deployment policy.

The City does not yet have enough publicly available information to justify a precise expansion number with confidence. A system handling more than 1.6 million medical incidents annually varies substantially by borough, hour, acuity, hospital destination and staffing condition. The same nominal number of ambulances can produce very different response reliability depending on where and when the units operate, how long they remain committed, how frequently tours are actually filled, and how often vehicles are unavailable.

A credible reform plan should therefore specify the method for determining capacity, the performance standards the system is expected to meet, and the data that will be publicly reported—not force the analysis to support a number chosen in advance.

From Diagnosis to Coordinated System Improvement

The diagnosis points to multiple constraints that interact with one another; no single intervention will ensure reliable EMS performance. The following actions should therefore be pursued as a coordinated improvement program, with some proceeding simultaneously. The first priority is to improve the performance and resilience of the entire EMS system before considering any specialized response overlay.

Reduce Avoidable Ambulance Demand Without Discouraging 911 Use

The objective is not fewer 911 calls. It is to ensure that each caller receives the most appropriate response while preserving ambulance availability for patients who require ambulance-level care. Safety must remain the governing principle. No caller should be discouraged from using 911 when an

emergency is suspected, and alternative pathways should be used only after the initial emergency screen and under defined medical oversight.

- **Secondary clinical triage:** use a second-level nurse, paramedic or physician assessment only for carefully selected low-acuity calls after the initial emergency screen, with defined exclusion criteria, immediate escalation to an ambulance when clinical uncertainty exists, and documented medical oversight.
- **Telehealth/treat-in-place:** expand FDNY/H+H telehealth only for validated low-acuity cohorts; track 24-, 48- and 72-hour repeat 911/ED use, adverse events, patient completion of alternative care and cases requiring later escalation.
- **Behavioral-health alternatives:** route appropriate nonviolent behavioral-health calls to specialized teams when safe and operationally appropriate.
- **Frequent-user intervention:** identify high-utilizing patients and coordinate EMS, primary care, behavioral health, social services, housing and payer resources to break repetitive 911–ED cycles.
- **Alternative destinations/non-emergency transportation:** evaluate clinically appropriate options where permitted, with clear exclusion criteria and medical oversight.
- **Institutional outreach:** identify nursing homes, shelters and other facilities generating unusually high potentially avoidable EMS demand and provide targeted clinical support and telemedicine access.

Employ demand-management metrics to continually evaluate effectiveness. Measure ambulance responses and ED transports safely avoided after clinical review, alternative-care completion, repeat 911/ED use at 24/48/72 hours, adverse events, later escalation to emergency care, patient experience and ambulance-hours returned to availability. Do not set a target of simply reducing 911 calls, and do not expand alternative pathways if safety outcomes deteriorate.

Recover Capacity Now Being Lost

Hospital turnaround. Reducing hospital turnaround should be pursued aggressively because every minute recovered returns ambulance-hours to the street. FDNY reports that Hospital Liaison Officers have reduced turnaround by at least seven minutes at participating hospitals.[5] This is a documented capacity-recovery opportunity, but the Task Force should separately quantify how much of the current response-time deterioration is attributable to hospital delay.

Hospital interface. Report ambulance patient-offload and hospital turnaround distributions by receiving facility, and determine how much operational capacity can be recovered through joint FDNY–hospital action.

Dynamic deployment. Ambulance availability is geographic and time-sensitive. Predictive deployment should incorporate current unit status, predicted 15–30 minute demand, pending calls, travel times and hospital conditions so available resources are positioned where they are most likely to be needed.

System resilience. Include emergency communications continuity and dispatch-system resilience in the operational assessment; reliability must be maintained during peak demand and system disruptions.

Stabilize the Workforce and Address Morale

Ambulance capacity is a workforce issue as much as a fleet issue. Recruitment, retention, vacancies, recurring overtime, morale and adequate compensation determine whether scheduled ambulance tours can actually be staffed.

The Comptroller reports more than 500 departures from EMT titles in FY2026 (19.5% of average EMT headcount). Among EMT and paramedic title departures, 58% and 45%, respectively, were movements to firefighter titles—not necessarily separations from City employment. Its approximately \$289 million FY2027 estimate for firefighter-level EMS compensation is a modeled scenario subject to assumptions, not a negotiated agreement; it should not be treated as interchangeable with other compensation estimates. The City should assess net staffed ambulance-hours lost through departures, transfers, vacancies and training delays. [9]

Why morale matters. Morale is an operational capacity issue when it affects turnover, absenteeism, recruitment, willingness to accept overtime and employees' decisions to build careers in EMS. Competitive pay, workload, facilities, equipment reliability, supervisory support, career opportunity and whether employees believe EMS is institutionally valued all influence retention.

- **Address competitive compensation and compensation equity:** treat compensation as a core recruitment and retention issue, while recognizing that specific salary changes and collective-bargaining terms belong in the OLR labor-relations process. Keep those costs transparent and separate from the cost of adding operational ambulance capacity.
- **Establish explicit EMT and paramedic vacancy targets** and track hires, academy completion, resignations, retirements and EMS-to-Fire promotions.
- **Evaluate direct/lateral entry for already-certified experienced paramedics** to supplement the current internal promotional pipeline.
- **Reduce chronic reliance on overtime** by converting recurring staffing needs into permanent positions; track overtime by employee and tour as a workforce-health indicator.
- **Conduct an EMS facility assessment** covering station condition, locker/shower/restroom and meal space, staging, restocking, training space, parking, maintenance backlog and geographic adequacy.

Determine, Fund and Build the Ambulance Capacity the Data Show Is Required

The City must fund more than staffed tours. Each new ambulance requires a serviceable vehicle, medical and communications equipment, a replacement plan and sufficient spare capacity so preventive maintenance or repairs do not cause funded tours to go unoperated.

Capacity definitions. Distinguish authorized or scheduled ambulance tours, actually staffed and operated tours, in-service ambulance-hours and units immediately available for assignment. The capacity model should identify lost hours attributable to vacancies, leave, hospital turnaround, maintenance and other causes before sizing new resources.

The Comptroller's fleet findings make it especially important to separate mechanical readiness, daily fleet-service targets, staffed in-service ambulance-hours, units committed to incidents and units actually available for immediate assignment. The Task Force should obtain unit-status and CAD data by borough, tour and demand period to establish the relationship between capacity and clinical performance. [9]

The City should determine the ambulance expansion required for reliable systemwide EMS performance empirically rather than begin with a fixed fleet number. FDNY and OMB should use incident-level CAD data and operational modeling to calculate the additional staffed ambulance-hours needed to meet defined reliability and clinical-response standards citywide and by borough, tour and

demand period. This Base EMS Capacity Requirement should be established independently of any later decision about a specialized Ultra-Time-Critical response.

FDNY should build a transparent model that first determines the staffed ambulance-hours, BLS/ALS mix, tours and geographic distribution required to meet adopted systemwide and borough-level reliability standards without relying on a specialized response overlay. The model should incorporate call volume and acuity by hour and geography; unit utilization; time from dispatch through return to service; transport rate; hospital turnaround by receiving facility; scheduled versus operated tours; ALS/BLS availability; travel-time distributions; projected vacancies; leave and relief factors; fleet downtime and spare ratio; simultaneous incidents; and pending-call backlogs. The resulting staffed hours should then be translated into staffing, physical ambulances, reserve vehicles, maintenance capacity and facilities.

Incremental CRT Partnering Requirement. After the Ultra-Time-Critical cohort and partnering standards are defined, FDNY should calculate only the additional ambulance capacity, if any, needed above the Base EMS Capacity Requirement to ensure reliable simultaneous CRT/ambulance dispatch. Some or all of the partnering requirement may already be satisfied by the base expansion; the model must prevent double-counting.

Capacity determination should be iterative. Initial resources should be placed where modeling identifies the largest reliability gaps, then the model should be recalibrated using actual results. Additional capacity should be added until agreed response-percentile, backlog, unit-availability and CRT-partnering standards are consistently achieved. The process should stop at the level of capacity the data justify, not at an arbitrary preselected ambulance count.

Decision scenarios. For OMB and City Hall review, the Task Force should present low, base and high capacity scenarios showing the staffing, ambulance-hours, fleet, facilities and recurring/capital costs associated with each level of performance. This preserves data-driven planning while giving decision-makers clear fiscal choices rather than an open-ended request.

- **Develop a capital acquisition plan** for the additional ambulances and their medical, communications and safety equipment. Any later purchase of specialized critical-response vehicles should occur only after the clinical model, coverage area and pilot requirements are defined.
- **Use FDNY's actual operating standard** to determine the required spare ratio; do not assume one physical vehicle is sufficient for each continuously operated tour.
- **Expand Fleet Services** staffing, parts inventory, preventive-maintenance capacity, road service and maintenance-bay capacity as required by the larger fleet.
- **Establish a lifecycle replacement schedule** based on age, mileage, engine hours, maintenance history and reliability.

For management purposes, vehicle downtime should be treated like a staffing vacancy: either condition produces one fewer ambulance available to answer calls.

Optional Ultra-Time-Critical Response: Define the Need Before Designing the Pilot

Only after the City has established and begun implementing the Base EMS Capacity Requirement should FDNY determine whether a specialized response is still needed for a narrow subset of emergencies in which exceptionally rapid intervention may materially affect survival or serious morbidity. If the performance data show such a gap, FDNY Medical Affairs should define the Ultra-Time-Critical Medical Emergency cohort, beginning with elements within existing Segment 1

emergencies such as cardiac arrest and choking and adding other presentations only where clinical evidence supports exceptionally rapid intervention. A Critical Response Team/Critical Response Vehicle (CRT/CRV) would then be considered as one possible response model: a specialized, non-transporting clinical resource dedicated to that cohort and always paired with a simultaneously dispatched, fully staffed ambulance.

If the CRT/CRV model is selected, FDNY and OMB should calculate an Incremental CRT Partnering Requirement separately from the Base EMS Capacity Requirement. The analysis should determine only the additional ambulance capacity, if any, required to assure simultaneous partner-ambulance availability without weakening ordinary EMS coverage. Some or all of that requirement may already be satisfied by the base system expansion, and the model should prevent double-counting.

The expense budget should fund EMTs, paramedics, supervisors, training, fringe benefits, medical supplies and recurring operating costs. The capital plan should separately fund ambulance acquisition and equipping, reserve fleet, station/staging improvements, major fleet-maintenance infrastructure and lifecycle replacement. Any specialized response vehicles should be funded only after FDNY Medical Affairs defines the clinical need, the Task Force determines that the option is operationally justified, and the required partner-ambulance capacity is established.

Existing FDNY Paramedic Response Units (PRUs): experience to evaluate

FDNY already operates nontransporting Paramedic Response Units. As described in the October 20 testimony, their broader role includes improving ALS availability by allowing paramedics to transfer appropriate noncritical patients to BLS ambulances and return to service. The proposed CRT/CRV concept differs in its narrowly defined ultra-time-critical cohort, limited geographic pilot, simultaneous transport-ambulance dispatch, explicit partnering safeguards and evaluation of clinical intervention and outcomes. CRT/CRV should be assessed as a possible complement to, not an assumed replacement for, PRUs.

An important operational concern remains unproven. When transport ambulances are scarce, a PRU arriving first may remain committed to a patient while awaiting transport, potentially delaying hospital care and reducing PRU availability for other calls. This is a plausible risk, not an established frequency or demonstrated clinical outcome. We have not identified public FDNY data sufficient to prove or disprove the relationship, and should not assume that FDNY has never analyzed it.

The Task Force should evaluate PRU dispatch and unit-status records: PRU-to-ambulance arrival intervals; frequency and duration of waits for transport; PRU scene and total commitment times; return-to-service after transfer; ALS availability gained; the relationship of these measures to ambulance availability, pending calls, geography, tour and acuity; and any association with clinical outcomes using appropriate risk adjustment. It should also determine whether staffing PRUs reduces ALS coverage elsewhere. Results should inform whether a CRT/CRV pilot is warranted and its dispatch, staffing and partner-ambulance standards.

The Comptroller's capacity and workforce findings increase the importance of testing this concern, but do not prove that PRU-to-transport delays have increased or caused worse patient outcomes. The Task Force should stratify PRU-to-ambulance intervals, PRU scene commitment and subsequent unit availability by actual transport-ambulance availability, pending-call backlog, borough and demand period, with appropriate clinical outcome measures and risk adjustment.

If Needed, Pilot the Ultra-Time-Critical Response Only After the System Foundation Is in Place

CFR-D role. FDNY CFR-D-trained fire personnel can provide early CPR, defibrillation and other appropriate first-response interventions on selected medical emergencies. Their arrival should count toward the six-minute clinical-response metric only when the crew is trained, equipped and authorized to provide the immediate intervention required by that call. For needs beyond CFR-D scope, the clock continues until an appropriately qualified CRT/CRV or ambulance arrives.

CRT staffing safeguard. Staffing the CRT/CRV must not reduce scheduled ALS ambulance capacity or worsen existing ambulance vacancies. CRT positions should come from incremental staffing or another arrangement demonstrated to be capacity-neutral. If the City cannot staff the pilot without cannibalizing ambulance tours, the pilot should not begin.

CRT/CRV pilot. If performance data support proceeding, the pilot should be the last operational step. It should begin only after the Base EMS Capacity Requirement is in sustained operation and any Incremental CRT Partnering Requirement has been funded and put in place. Each pilot CRT/CRV should initially cover a deliberately limited geographic area with a sufficient concentration of eligible incidents to support reliably faster arrival. CAD data should identify the initial area using incident density, time-of-day patterns, travel-time distributions and partner-ambulance availability. The geographic coverage area should expand only when measured results show that the response and partnering standards can continue to be achieved without weakening general EMS coverage.

Proposed pilot standards. Subject to validation by FDNY Medical Affairs, the pilot should test a 90% within six minutes goal for the first clinically capable responder, a 90% within eight minutes goal for the fully staffed ambulance, and—when the CRT/CRV arrives first—a 90% within three additional minutes goal for the partner ambulance.

Why the three-minute gap. The three-minute ambulance-arrival interval is an operational safeguard, not a claimed clinical breakpoint. It is intended to ensure that an early-treatment resource functions as an extension of a complete EMS response rather than as a stand-alone unit left waiting for transport capability. Persistent gaps beyond three minutes—and especially beyond five—should trigger review of ambulance capacity, dispatch practices and the CRT coverage area rather than expansion of the pilot.

The intended sequence: 911 recognition → ambulance plus closest clinically capable responder dispatched together → early lifesaving treatment → continuing ambulance care and transport when clinically indicated.

The sequence to avoid: CRT arrives → assesses patient → requests ambulance → waits.

Safeguard: No CRV without a fully staffed ambulance behind it.

The three-minute partner-ambulance interval in the original proposal is a proposed operational safeguard after CRT arrival, not a guarantee of three-minute CRT arrival. The testimony also discusses an ambitious approximately three-minute CRT arrival objective within a deliberately limited pilot area. These are different clocks. Both require clinical and operational validation; neither should be described as a demonstrated citywide standard.

Public Reporting, Clinical Outcomes and Operational Reliability

The initiative should be judged by patient-oriented performance and operational reliability, with results made public. Averages alone can conceal long-tail delays and geographic disparities. FDNY should report the percentage of incidents meeting clinically appropriate, condition-specific standards; median, 75th and 95th percentile response times; and the component intervals of the full 911-to-arrival process. Reports should be stratified by borough, tour and demand period, and by community board where statistically reliable. A monthly executive scorecard and downloadable data should be supplemented by periodic appropriately risk-adjusted clinical-outcome evaluation.

- **Response and backlog transparency:** publish 50th, 75th and 95th percentile end-to-end response times; percentage meeting clinically appropriate standards; defined pending-dispatch counts and duration; and the number of calls exceeding relevant delay thresholds. Introduction 990-2026 proposes notification when more than 25 ambulance dispatches are pending because of insufficient available units; the Task Force should examine this indicator and appropriate notification criteria regardless of the bill's disposition, without discouraging necessary 911 calls.

Publish reconciliation of official response-time series and their denominators; distinguish fleet-service readiness from staffed and dispatchable capacity; and report the frequency, duration, location and clinical priority of dispatch backlogs, including periods with more than 25 pending ambulance assignments. These are diagnostic measures, not substitutes for patient outcomes. [9]

- **Ultra-Time-Critical clinical performance:** first clinically capable responder type and arrival time; ambulance arrival; CRT/CRV-to-ambulance gap; time to first defibrillation where applicable; return of spontaneous circulation; airway/hemorrhage interventions; survival to hospital admission and discharge where valid and available; and neurologically intact survival for appropriate cardiac-arrest cohorts. Evaluation should use appropriate risk adjustment and, where feasible, comparison areas or pre/post controls so changes are not attributed to the pilot when they are driven by case mix, bystander CPR, hospital care or other external factors.
- **Workforce:** authorized, budgeted and actual EMT/paramedic headcount; vacancies by title; hires; academy entry/completion; resignations; retirements; EMS-to-Firefighter promotions; overtime; sick/limited duty; and turnover by years of service.
- **Capacity and fleet:** scheduled versus operated ambulance tours; in-service ambulance-hours; staffed ambulance-hours by borough and tour; ALS/BLS shortfalls; unit utilization/busy fraction; peak ambulances; fleet availability; vehicles out of service; maintenance downtime; spare ratio; preventive-maintenance compliance; and the modeled versus actual ambulance-hours required to meet adopted response and partnering standards.
- **Hospitals and demand management:** turnaround at the 50th, 75th, 90th and 95th percentiles by hospital; ambulance responses and ED transports safely avoided after clinical review; alternative-care completion; 24-/48-/72-hour repeat 911 or ED use; adverse events; and year-to-date medical incident volume.

Task Force Recommendations and Implementation Milestones

The Mayor-directed EMS Improvement and Modernization Task Force should turn the proposed system improvements into a coordinated, measurable implementation program. Several workstreams—workforce stabilization, hospital turnaround, demand management, dispatch reliability and capacity analysis—should proceed in parallel. The following milestones identify the deliverables, responsible partners and decisions expected at each stage.

Within 30 days — Organize and establish the baseline. Convene senior representatives from FDNY, OMB, OLR, DCAS, NYC Health + Hospitals, GNYHA, EMS labor and other necessary partners; designate an accountable lead for each workstream; establish data-sharing protocols; agree on response-clock definitions, clinical categories and geographic reporting; and adopt an initial dashboard for response times, available staffed ambulance-hours, pending dispatches, workforce and hospital turnaround.

Within 60 days — Complete the initial system assessment and begin immediate remedies. Reconcile differing official response-time series and fleet-readiness measures; measure scheduled, staffed, in-service and immediately dispatchable ambulance-hours by borough and tour; quantify vacancies, attrition, EMS-to-Firefighter movement, overtime, fleet downtime, dispatch delays and hospital patient-transfer delays; map demand, backlogs and geographic coverage gaps; assess existing PRU performance and transport-ambulance partnering; and identify actions that can begin under existing authority without waiting for the full plan. The Comptroller's October 16 report should be considered as one source of evidence in this assessment, alongside FDNY operational records and other City and hospital data.

Within 90 days — Submit the implementation and funding plan. Present clinically appropriate performance targets and the percentage of calls expected to meet them; calculate the Base EMS Capacity Requirement using demand, utilization, coverage and staffed ambulance-hours; specify the operational changes needed to recover existing capacity; set out workforce and compensation/retention options; assign owners and deadlines; and provide low, base and high cost scenarios with separate expense and capital requirements. Define any Incremental CRT Partnering Requirement separately, without double-counting transport ambulance capacity. Submit the plan to the Mayor and Council for review.

Within six months — Implement and test. Put priority hospital-turnaround, workforce, fleet-uptime, demand-management and predictive-deployment measures into operation; track whether they recover staffed and dispatchable ambulance-hours and improve response reliability. Only if clinical and operational evidence supports it, begin a geographically limited CRT/CRV pilot with dedicated additional staffing, simultaneous transport-ambulance dispatch, explicit ambulance-partner standards and protection of existing ALS capacity.

Within twelve months — Evaluate outcomes and revise the program. Publish results against baseline and targets, including median, 75th and 95th percentile response times, percentages meeting clinically appropriate standards, borough and community-level reliability where statistically sound, staffed and available ambulance-hours, pending dispatches, hospital turnaround, workforce retention, fleet availability and selected risk-adjusted patient outcomes. Evaluate pilots using comparison areas or pre/post controls where feasible, then continue, modify or discontinue interventions based on evidence.

Ongoing accountability. Issue a monthly public executive scorecard and downloadable operational measures, with periodic clinical-outcome reporting and Council oversight. The Task Force should report obstacles requiring interagency or budget action and update implementation priorities as results become available.

A Coordinated Path Forward

Reduce avoidable demand. Recover lost capacity. Stabilize the workforce and morale. Use verified operational data to distinguish documented conditions from factors that still require quantification.

Calculate and fund the Base EMS Capacity Requirement. Measure whether systemwide reliability and clinical performance improve. Only if the data demonstrate a remaining need for a specialized Ultra-Time-Critical response should the City design and test that option, with any incremental partner-ambulance capacity calculated separately. Report the assumptions, operational measures and risk-adjusted clinical results publicly so the City can revise the system as conditions change.

Source Notes

[1] NYC Preliminary Mayor's Management Report FY2026 and NYC Council FY2027 FDNY budget materials: FY2025 incident volume, life-threatening incident volume, FY2026 in-service ambulance-hours, response-time trend and related staffing/budget measures.

[2] CBS News New York, Aug. 20, 2026, reporting FDNY July 2026 average response of 14:40 for life-threatening medical emergencies and proposed City Council backlog-notification legislation.

[3] FDNY, Feb. 2026 probationary firefighter graduation announcement: 126 of 181 graduates had previously served as EMS officers, paramedics or EMTs.

[4] FDNY/JoinFDNY EMS career materials: Paramedic is described as a promotional opportunity available to FDNY EMS members; EMTs may pursue the promotional Firefighter pathway.

[5] FDNY 2025 End-of-Year Accomplishments: Hospital Liaison Officer and telehealth initiatives referenced in the capacity-recovery and demand-management sections.

[6] NYC Open Data, EMS Incident Dispatch Data: recommended source for incident-level CAD analysis and validation of demand, geography, travel time and backlog patterns.

[7] James T. Kerr, Consolidated Written and Oral Testimony for the New York City Council Committees on Fire and Emergency Management and Hospitals, October 20, 2026 (version dated October 17, 2026), including FY2026 Mayor's Management Report and September 2026 Local Law 119 data, and discussion of Introduction 990-2026.

[8] FDNY PRU program description as summarized in the October 15, 2026 Council testimony. Proposed analysis of PRU-to-ambulance delays is a research recommendation, not a finding from verified PRU incident-level data.

[9] Office of the New York City Comptroller, A System Under Strain: FDNY EMS Operations and Fiscal Pressures (October 16, 2026), <https://comptroller.nyc.gov/reports/a-system-under-strain-fdny-ems-operations-and-fiscal-pressures/>. Figures and qualifications are reflected in the October 17, 2026 Council testimony revised to incorporate the Comptroller's findings.