nature human behaviour

Article

https://doi.org/10.1038/s41562-024-01889-6

A randomized controlled trial evaluating the effects of nurse-led triage of 911 calls

Received: 22 July 2023

Accepted: 16 April 2024

Published online: 24 May 2024

Check for updates

Kevin H. Wilson [1,5] ✉, Rebecca A. Johnson [1,2,5], Chrysanthi Hatzimasoura[1], Robert P. Holman[3], Ryan T. Moore [1,4] & David Yokum[1]

To better connect non-emergent 911 callers to appropriate care, Washington, DC, routed low-acuity callers to nurses. Nurses could provide non-emergent transportation to a health centre, recommend self-care or return callers to the traditional 911 system. Over about one year, 6,053 callers were randomized (1:1) to receive a business-as-usual response ($n_{control}$ = 3,023) or further triage ($n_{treatment}$ = 3,030). We report on seven of nine outcomes, which were pre-registered (https://osf.io/xderw). The proportion of calls resulting in an ambulance dispatch dropped from 97% to 56% (β = −1.216 (−1.324, −1.108), P < 0.001), and those resulting in an ambulance transport dropped from 73% to 45% (β = −3.376 (−3.615, −3.137), P < 0.001). Among those callers who were Medicaid beneficiaries, within 24 hours, the proportion of calls resulting in an emergency department visit for issues classified as non-emergent or primary care physician (PCP) treatable dropped from 29.5% to 25.1% (β = −0.230 (−0.391, −0.069), P < 0.001), and the proportion resulting in the caller visiting a PCP rose from 2.5% to 8.2% (β = 1.252 (0.889, 1.615), P < 0.001). Over the longer time span of six months, we failed to detect evidence of impacts on emergency department visits, PCP visits or Medicaid expenditures. From a safety perspective, 13 callers randomized to treatment were eventually diagnosed with a time-sensitive illness, all of whom were quickly triaged to an ambulance response. These short-term effects suggest that nurse-led triage of non-emergent calls can safely connect callers to more appropriate, timely care.

There are an estimated 240 million 911 calls in the United States every year—about 460 every minute on average[1]—and approximately one in ten of these is directed to emergency medical services (EMS). Since the 911 call system was launched in 1968, a variety of approaches have been attempted to triage how EMS are deployed. These approaches tend to focus on reserving relatively scarce and costly advanced life support units—lights-and-sirens ambulances, for example—and emergency department (ED) capacity for use in high-acuity emergencies, such as a heart attack or gunshot wound. One tactic, based on research finding that many people were relying on 911 and EDs as a means to access health care they otherwise could not access (for example, due to lack of insurance or a primary care physician (PCP)), has been to expand access (for example, through Medicaid coverage expansions and promotion of primary care resources in Federally Qualified Health Centers (FQHCs))[2].

Many studies suggest, however, that EMS systems continue to face strain from calls for non-emergent health issues such as common cold symptoms and ankle sprains. EDs also remain overutilized, with research estimating that about half of ED visits could be avoided, with a potential savings of US$38 billion per year[3]. However, there is understandable concern about inadvertently denying needed care

[1]The Lab @ DC, Washington, DC, USA. [2]McCourt School of Public Policy, Georgetown University, Washington, DC, USA. [3]Fire and Emergency Medical Services, Washington, DC, USA. [4]School of Public Affairs, American University, Washington, DC, USA. [5]These authors contributed equally: Kevin H. Wilson, Rebecca A. Johnson. ✉e-mail: thelab@dc.gov

https://doi.org/10.1038/s41562-024-01889-6

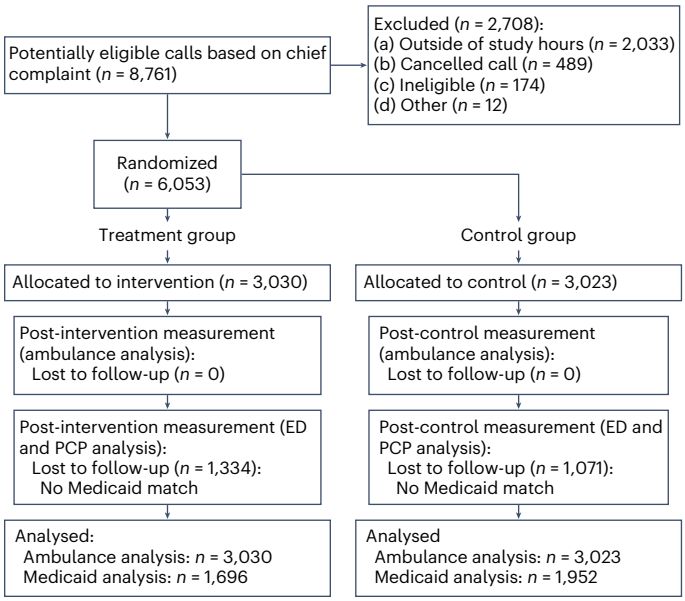


Fig. 1 | CONSORT diagram. The progression through the RCT to reach randomization and measurement of outcomes.

to any particular 911 caller. But there is also widespread concern that many jurisdictions have overwhelmed emergency response capacity, meaning that, system-wide, many emergency callers might be at risk of insufficient care. There remains, in other words, a need to more effectively triage calls, but without unduly restricting needed patient care.

Our paper describes a randomized controlled trial (RCT) evaluating a new form of triage within the 911 call system of Washington, DC. The programme—called 'Right Care, Right Now', or the Nurse Triage Line (NTL)—identified a subset of calls as low acuity and appropriate for triage to more appropriate care than a normal ambulatory response. Calls in this subset were randomly assigned to either a business-as-usual response from call takers or a warm hand-off to a nurse who identified the most appropriate form of care, which could be an ambulance but could also be free transport to a primary care facility or care over the phone. As elaborated below, for 10.5 months starting in April 2018, a subset of 6,053 callers to 911 were randomly assigned to talk with a nurse before a decision was made about whether to send EMS. This nurse, importantly, was also empowered to redirect the caller towards a same-day appointment, with free transportation via cab or rideshare, at a nearby clinic for primary care. The theory of impact was that the nurse could help avoid EMS deployments and potentially even improve health outcomes by diverting the caller into better-matched primary care rather than the more limited treatments available from an ambulance or ED.

Nurses have helped triage in a variety of times and places, ranging from battlefield care during World Wars I and II to, more recently, filtering whether and how patients are admitted into EDs[4,5]. Telephone hotlines staffed by nurses have also helped callers obtain consultative services outside of in-person clinic hours or, notably for present purposes, provided an alternative to calling 911. For example, a community health care initiative by the Regional Emergency Medical Services Authority in Reno, Nevada, included a 24/7 Nurse Health Line, a ten-digit telephone number residents could call instead of 911. In 2017, the Regional Emergency Medical Services Authority reported that their nurse line and alternative transports had saved over 4,400 ED visits and prevented over 630 transports between October 2013 and June 2016 (ref. 6). At least two other US jurisdictions (MedStar in Fort Worth, Texas, and LMEMS in Louisville, Kentucky) have embedded nurses within their 911 call systems, similar to the Washington, DC, pilot we report here. A non-randomized, retrospective cohort study found that the Medstar and LMEMS programmes reduced ambulance transports and increased self-reported patient satisfaction[7]. Outside of the United States, observational studies of calls to emergency systems show that many calls are for issues that do not require an immediate lights-and-sirens emergency response[8]. In these non-US contexts, pre–post evaluations show that non-emergency helplines can receive substantial uptake even if they require calling a separate number[9], and, similar to the US results, that triage is associated with reductions in ambulance dispatches and visits to the ED[10].

Our project further builds the evidence base about 911 triage in three ways. First, we implement the concept in a different jurisdiction and provide another demonstration of its operational feasibility. Second, we measure not only EMS deployments but also impacts on primary care utilization, which is a preliminary window into whether 911 triage provides not only triage but also diversion; that is, does the intervention simply lower the chance of EMS deployment or can it also increase the likelihood a person goes to a clinic instead? Third, we provide experimental evidence of the effectiveness of triage within a 911 system. While analyses of what happens before and after a triage system are important (such as those in Fort Worth and Louisville), they fail to isolate the causal effect of the new system from other changes that might happen simultaneous to the reform—for instance, changes to the dispatch system, temporal trends in 911 calling, and other factors. Our call-level randomization provides better-identified causal estimates of the triage reform's effects.

We note that Washington, DC, is an ideal setting for two reasons. The first is the scale of its 911 response system and the strains on that system: DC has one of the highest per capita volumes of 911 medical responses in the country and estimates that up to one in four callers does not require emergency services[11]. Second, due to DC's early Medicaid expansion under the Patient Protection and Affordable Care Act, DC has one of the nation's highest rates of Medicaid coverage among those eligible, which translates into near-universal health insurance coverage[12]. Thus, 911 callers are extremely likely to have subsidized access to non-emergency care options, allowing DC to design a diversionary intervention without worrying about whether callers can pay. Additionally, since DC has direct access to Medicaid claims data, the high proportion of Medicaid enrolment allows us to observe total health care utilization for a relatively high proportion of callers, allowing us to contribute to the literature on how Medicaid coverage relates to ED and primary care utilization[13–16].

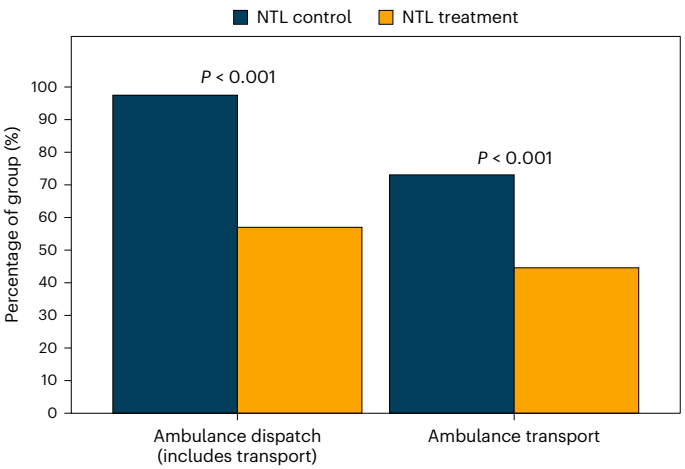


Fig. 2 | Percentages of the control and treatment groups receiving ambulance dispatch and transport. The *P* values in the figure are from a logistic regression of each outcome on the treatment indicator and a two-tailed test of whether the coefficient is equal to zero. The exact *P* values and test statistics are provided in Table 1.

https://doi.org/10.1038/s41562-024-01889-6

Table 1 | Effect of the NTL on ambulance dispatches and transport

	Dependent variable		
	Ambulance transport (last status)	Ambulance transport (any status)	Ambulance dispatch (any status)
Treatment	−1.216	−1.262	−3.376
	(−1.323, −1.108), $P < 0.001$	(−1.371, −1.153), $P < 0.001$	(−3.615, −3.137), $P < 0.001$
Constant	0.998	1.102	3.658
	(0.918, 1.079), $P < 0.001$	(1.019, 1.184), $P < 0.001$	(3.430, 3.886), $P < 0.001$
No. of observations	6,053	6,053	6,053
Log likelihood	−3,843.000	−3,789.000	−2,425.000
AIC	7,691.000	7,581.000	4,855.000

The results are from a logistic regression of each ambulance use measure on the treatment status, presenting the β estimate and 95% confidence interval. The P values correspond to a two-tailed test of the null hypothesis that the coefficient is equal to zero and are not adjusted for multiple comparisons. The exact P values are 1.5×10^{-108} for ambulance use (last status), 7.3×10^{-114} for ambulance use (any status) and 5.2×10^{-169} for ambulance dispatch. AIC, Akaike information criterion.

In the spring of 2018, DC began a US$1 million pilot of the nurse triage programme. The programme aimed to move low-acuity calls out of the emergency medical system and into the primary care system, and thus was designed to provide alternative and medically responsible care for DC's non-critical, yet time-sensitive, 911 callers by connecting them to primary and urgent care providers better positioned to deliver care. Furthermore, behavioural science teaches that process design changes should remove as many points of friction as possible, even small ones[17], so we placed nurses directly in the 911 dispatch loop instead of creating a separate nurse helpline.

Our research team worked with DC's Fire and Emergency Medical Services Department (FEMS) to incorporate a randomized evaluation of this programme into the rollout. Here we report on seven pre-registered primary outcomes: changes in the rate of (1) ambulance dispatches and (2) transports from 911 calls, changes in the rates of Medicaid beneficiaries' visits to PCPs within (3) 24 hours and (4) six months of interacting with the nurse help line, changes in the rate at which Medicaid beneficiaries subsequently received a diagnosis within (5) 24 hours and (6) six months from an ED where the diagnosis's ICD-10 code indicates that it was probably 'non-emergent' or 'primary-care treatable' and (7) total log-transformed Medicaid expenditures within six months of a 911 call. Two of our pre-registered primary outcomes—measuring the rate of repeat callers to the 911 system between the control and treatment groups over two different periods—were ultimately infeasible due to insufficient data collected at the time of the call. Finally, we also report on the results of an exploratory patient satisfaction survey and an in-depth safety analysis.

Results

Between 19 April 2018 and 1 March 2019, 6,053 callers entered the randomized evaluation of DC's 911 nurse triage programme on the basis of an initial assessment of low acuity. Of these, 3,030 were assigned to the nurse (treatment group) and 3,023 were assigned to business-as-usual (control group). Figure 1 shows the participant flow. Because the point of randomization was the 911 call, we could not collect demographic information for all callers. Extended Data Fig. 1 shows the monthly count of calls—we see a sharp increase when the triage line extended its hours of operation from 7:00–23:00 to 24 hours during the second window.

Callers randomized to speak to a nurse were consistently less likely to receive an ambulance for their low-acuity issue than callers randomized to business-as-usual. Figure 2 shows the raw proportions and the P values from a logistic regression of receiving an ambulance dispatch or transport (labelled 'Ambulance transport') on a binary indicator for the treatment, with a two-tailed test to investigate whether the coefficient was equal to zero. The results show that the treatment caused a significant reduction in both whether an ambulance was dispatched at all (left), which includes both dispatches that stop at the dispatch stage and dispatches that result in a transport, and whether an ambulance was dispatched and resulted in an ambulance transport (right). In total, 97% of the control group callers had an ambulance dispatched compared with 56% of the treatment group (a 42% relative reduction; $\beta = -1.216$ (−1.323, −1.108), $P < 0.001$). Focusing on dispatches that resulted in cost-intensive transports, while 73% of the control group callers received an ambulance transport for their non-emergent or PCP-treatable issue, only 45% of the treatment group callers received an ambulance transport for their similar issue (a 38% relative reduction; $\beta = -3.376$ (−3.615, −3.137), $P < 0.001$). Table 1 shows additional results that separate ambulance dispatches into those resulting in a transport and those that did not, and Extended Data Table 1 shows that the treatment effect is large and statistically significant in each of the study months.

For the Medicaid beneficiaries, we examined the programme's impact on health care utilization. These beneficiaries comprise 3,067 patients, or 50.7% of the sample. Supplementary Information section 2 describes the matching process and Medicaid beneficiary characteristics—we see that whether a respondent has enough identifiers to be matched is related to their ambulance transport status (Extended Data Table 2), but we see similar characteristics across callers randomized to the treatment and control groups (Extended Data Fig. 2 and Extended Data Table 3) and general characteristics (75% Black; average age ~45 years old) that resemble the broader DC Medicaid population. Moving to the analytic results, Fig. 3 summarizes the impact on visits to PCPs, which the programme aimed to increase, in the 24 hours and six months following a call. The figure shows significant increases in visits in the 24 hours post-call (going from 2.5% of callers visiting to 8.2% of callers visiting a PCP, a 228% increase; $\beta = 1.252$ (0.889, 1.615), $P < 0.001$). However, we fail to reject the null hypothesis of no differences over the longer period, as rates increase to closer to 43% of patients from each group (see Extended Data Table 4 for the exact percentages and Table 2 for the regression results).

Figure 3 shows similar patterns for visits to the ED for non-emergent or PCP-treatable issues. In the short term, while 29.5% of control group callers went to the ED for a non-emergent or

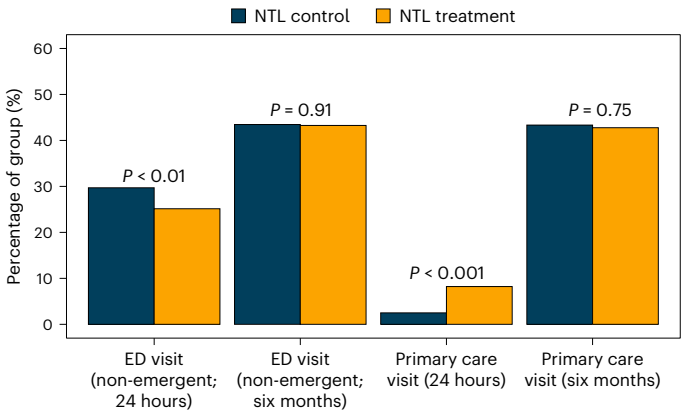


Fig. 3 | Percentages of the control and treatment groups visiting an ED or PCP. We examined ED visits for conditions that were classified as non-emergent or PCP treatable and all PCP visits. These were measured over two time spans: the 24 hours immediately following the call and the six months following the call. The P values in the figure are from a logistic regression of each outcome on the treatment indicator and a two-tailed test of whether the coefficient is equal to zero. The exact P values and test statistics are provided in Table 2.

https://doi.org/10.1038/s41562-024-01889-6

Table 2 | Effect of the NTL on ED visits for non-emergent or PCP-treatable conditions and PCP visits

	Dependent variable			
	24 hours post-call	Six months post-call	24 hours post-call	Six months post-call
	Non-emergent or PCP-treatable ED visit		PCP visit	
Treatment	−0.230	−0.008	1.252	−0.023
	(−0.389, −0.070), P=0.005**	(−0.151, 0.135), P=0.912	(0.889, 1.615), P<0.001	(−0.167, 0.120), P=0.752
Constant	−0.862	−0.263	−3.665	−0.268
	(−0.967, −0.756), P<0.001	(−0.361, −0.166), P<0.001	(−3.975, −3.355), P<0.001	(−0.366, −0.171), P<0.001
No. of observations	3,067	3,067	3,067	3,067
Log likelihood	−1,802.000	−2,099.000	−596.200	−2,096.000
AIC	3,609.000	4,202.000	1,196.000	4,197.000

The results are from a logistic regression of each visit type on the treatment status, presenting the β estimate and 95% confidence interval. The P values correspond to a two-tailed test of the null hypothesis that the coefficient is equal to zero and are not adjusted for multiple comparisons. **P<0.01.

Table 3 | Effect of the NTL on Medicaid expenditures

	Dependent variable	
	Logged expenditure within 24 hours	Logged expenditure within six months
Treatment	−0.167	−0.079
	(−0.379, 0.045), P=0.123	(−0.327, 0.169), P=0.534
Constant	3.279	6.022
	(3.135, 3.424), P<0.001	(5.853, 6.191), P<0.001
No. of observations	3,067	3,067
R^2	0.001	0.0001
Adjusted R^2	0.0005	−0.0002
Residual s.e. (d.f.=3,065)	2.989	3.497
F statistic (d.f.=1; 3,065)	2.383	0.388

The results are from a linear regression of log-transformed total Medicaid expenditures on the treatment status, presenting the β estimate and 95% confidence interval. The P values correspond to a two-tailed test of the null hypothesis that the coefficient is equal to zero and are not adjusted for multiple comparisons.

PCP-treatable issue, only 25.1% of treatment group callers did the same (a 15% decrease; β = −0.230 (−0.391, −0.069), P < 0.001). In the long term, the rates converged at 43% for each group, and we fail to reject the null hypothesis of no differences.

Table 3 contains the results comparing log-transformed Medicaid expenditures. We fail to reject the null hypothesis of no difference between the control and treatment groups (β = −0.079 (−0.327, 0.169), P > 0.1) in the six-month window. Table 3 also includes the results of a comparison within a 24-hour window, where we also fail to reject the null hypothesis of no differences.

While we fail to reject the null hypothesis of no differences in Medicaid expenditures, we note that such expenditures are only a small part of the overall EMS funding landscape. While a full cost analysis is out of scope for this report, our findings above imply that roughly 900 ambulance rides were avoided, saving approximately 186,000 minutes of ambulance availability. DC's largest Medicaid Managed Care Organization estimates that the total savings between a lights-and-sirens response to a 911 call and a visit to a clinic via taxi are approximately US$850, implying at least US$764,000 in savings over the 10.5 months of the experiment.

Finally, we note that 218 callers responded to our patient satisfaction survey. Of these, 204 (93.5% (90.3%, 96.8%)) indicated that they were "satisfied" or "highly satisfied" with the service they had received. In addition to these patient assessments, Supplementary Information section 1 describes the in-depth results of a safety analysis performed on all calls, and Extended Data Fig. 3 summarizes these results. The safety analysis found that the nurse routed 5% of calls to advanced life support but that these routings did not contribute to delays in patient receipt of care.

Discussion

We estimate via a randomized controlled design that the nurse-led 911 triage programme in Washington, DC, caused a statistically significant reduction in EMS utilization within the first 24 hours of a call. The ambulance dispatch rate dropped from 97% to 56% (and the ambulance transport rate from 73% to 45%), and the rate of ED visits measured among Medicaid beneficiaries dropped from 30% to 25%. The difference in magnitude between these two reductions could stem from multiple factors. First, callers randomized to the control group could still be referred by a nurse to an ED and could either take an ambulance there or transport themselves to the ED. Second, the ED visits are measured among the subset of callers who are Medicaid beneficiaries, who may have higher baseline rates of ED utilization than callers not receiving Medicaid. Importantly, these reductions were achieved without measurable side effects—there was, in particular, no evidence that any 911 caller diverted away from an ambulance was later found to have needed an ambulance, and at most 7 of 3,182 (0.2%) callers experienced a delay in care longer than two minutes. A formal cost–benefit analysis is outside the scope of this report, but these findings are almost certainly of substantial policy relevance, with very rough estimates indicating that the intervention was at least cost-neutral.

Moreover, a significant number of callers were diverted towards alternative destinations within the first 24 hours of a call: the percentage of calls resulting in a visit to a primary care clinic within 24 hours more than tripled, from 2.5% to 8.2%. The call-level randomization found similar improvements as previous before–after observational studies—reductions in ambulance transports (the largest improvements) and reductions in ED visits—while providing new findings that these changes in utilization of emergency services are accompanied by increases in the utilization of non-emergency care and showing that the results for other outcomes are robust to an RCT that controls for confounding time trends.

NTL stakeholders were hopeful that the programme not only would aid EMS triage but also might serve a type of equitable diversion function, migrating a request for one type of service into the provision of a different, more appropriate service. Low-acuity 911 callers might not need EMS, but they do usually have real health care needs, and so the nurse could help them access the appropriate care more quickly—the 'right care, right now'—which in turn could lead to better health

outcomes than if they had received an ambulance and gone to the ED. Our current analyses do not estimate health outcomes, so we cannot test this hypothesis directly. However, we did observe an increase in primary care uptake, which other research has found to be correlated with better health outcomes[18].

The system had profound practical effects during the trial, and we expect even larger effects as the system matures. With the large treatment effect of a 28-percentage-point reduction in ambulance transports for non-emergency issues during the RCT, we avoided roughly 900 ambulance transports in an already overburdened system. Going forward, the agency and legislators have documented larger impacts on the system due to the continuation of the expansion to being available 24 hours per day and due to dispatchers becoming more comfortable routing low-acuity calls to the triage system. Recent statistics show that approximately 9.2% of emergency calls were sent to the triage line, representing over 14,263 calls to 911. Of these triaged calls, 43% were diverted to non-emergency destinations. Legislators who discuss emergency care with constituents in their area describe the line as a "vital investment to support residents' needs immediately and ease the strain on our ambulance and hospital system"[19].The evidence generated from our smaller-scale RCT has supported continued investment in the service for these larger, longer-term impacts.

Per our registered pre-analysis plan, we also report on ED and primary care use six months after the 911 call. Our motivation was to probe whether the 911 nurse interaction might set in motion a larger chain of positive impacts. For example, rather than a person later visiting the ED multiple times for uncontrolled diabetes, the intervention's diversion of this patient to a primary care provider in their first call could reduce the need for later utilization of the emergency system for diabetes management and co-morbidities. Ultimately, although we estimated the impact over this six-month time horizon, the interpretation of these estimates is complicated by fundamental limitations that arise from randomizing at the call rather than patient level. First, call-level randomization means that we might have two callers initially randomized to the treatment condition whose trajectories then diverge. Caller A might continue calling, receiving a mix of continued nurse triage and business-as-usual care. Caller B might stop after that first call, receiving a single dose of the treatment. The six-month analyses aggregate over these distinct types of callers. Second, the nature of identifier collection in the 911 call system (phone numbers but incomplete names and dates of birth) means that we cannot confidently separate callers into different groups (for example, single dose of treatment, single dose of control, mixed treatment and control over time, many doses of treatment and so on). This means that we cannot perform a robustness check for the six-month analysis that restricts to one-time callers to avoid challenges with people's treatment being diluted by subsequent randomizations to the control group.

While these first interpretation challenges of the six-month outcomes relate to call-level versus caller-level randomization, the longer time window also means that the binary 'ever visit' measures might conceal variation in the underlying counts. While there are slightly lower rates of at least one ED visit in the treatment group (43.2% compared with 43.5% in the control group) and slightly higher rates of at least one primary care visit in the treatment group (43.3% compared with 42.8% in the control group), we fail to reject the null hypothesis of no difference in rates. The binary count method, however, could mask marginal effects. For instance, if the treatment group averaged only one ED visit, but the control group averaged three ED visits, the 'yes visited once' measure might conceal these count-based differences. One piece of evidence against this interpretation, however, is that we also fail to reject the null of no differences in Medicaid expenditures, which aggregate across all visits rather than simplifying these into a binary measure. Due to challenges with many different ways to define what counts as a 'discrete visit' in claims data where procedures are delivered and charged over time, we focus on the binary measures in the present paper.

In sum, while the results of the intervention are both statistically significant and large in magnitude in the 24 hours following the call, there are several challenges in interpreting our failure to detect statistically significant changes over the six-month time horizon. Future research set up to estimate caller-level effects in addition to call-level effects—for example, by only randomizing callers at a point when reliable name and date-of-birth information has been provided—can help refine estimates of long-term impacts.

Our study found that a large metropolitan jurisdiction was able to implement a triage line within its 911 call system and substantially reduce the number of ambulance deployments and ED visits within 24 hours, with the safety analysis showing that the intervention also avoided any undue restriction to needed patient care. These findings add to a small but emerging literature suggesting that triage is potentially an effective reform in 911 call systems. Should other jurisdictions decide to introduce or expand NTL programming, we would encourage them to include evaluation as a part of the pilot. Multiple evaluated NTL programmes, in a variety of contexts, will empower more robust learning about how to optimize an NTL programme and sharpen estimates of causal impacts. A collection of evaluations would inform meta-analyses and cost–benefit analyses necessary to establish a firm evidence base about the NTL model and, in turn, inform policy debates about whether and how to fund and operate such programmes.

Methods

The study protocol was reviewed by the Western Institutional Review Board (WIRB Work Order No. 1-1033362-1), which decided that the "project does not involve [human subjects] research" and instead constituted a "quality-improvement activity". Explicitly, "the HHS regulations for the protection of human subjects do not apply to such quality improvement activities, and there is no requirement under these regulations for such activities to undergo review by an institutional review board, or for these activities to be conducted with provider or patient informed consent". The study was always planned as an RCT, and the board reviewed the study in its RCT form. Prior to fielding the RCT, the study was pre-registered with the Open Science Foundation (a common hosting site in the social sciences), with the identifier https://doi.org/10.17605/OSF.IO/T7NHJ, on 17 April 2018. This experiment is also retroactively registered with the ClinicalTrials.gov identifier NCT05589168 for the medical community on 21 October 2022. Supplementary Information section 3 discusses departures from the pre-analysis plan. The analyses were conducted in Python (v.3.8.6)[20] and R (v.4.1.2)[21].

Randomization and eligibility criteria

When a 911 call was placed, a call taker would perform an initial assessment of the caller's chief complaint using the Criteria-Based Dispatch system developed by King County, Washington (2013). Calls were enrolled in the study if they met all of three eligibility criteria:

1. The caller was a non-incarcerated adult calling on behalf of themselves (a 'first-party caller').
2. Call takers in the 911 dispatch system, on the basis of information presented in the call, classified the call as belonging to one of ~48 chief complaints. These include issues such as allergic reactions with no breathing difficulties, migraines or panic attacks. The list was developed by a set of medical experts who considered the capabilities of nurses to match callers with these issues to appropriate care.
3. A nurse was available.

For criterion two, the full initial list of eligible complaints appears in Appendix B of our pre-analysis plan. This set of complaints was modified on 21 June 2018 to eliminate one determinate code and combine two to drop the number of codes from 48 to 46. This decision was approved by both the 911 call centre director K.

Holmes and R.P.H. For criterion three, initially, nurses were available between 7:00 and 23:00, but this was expanded to 24-hour operation in January 2019. As discussed in 'Power analysis', the trial was initially supposed to run from mid-April 2018 through mid-October 2018, but it was extended on the basis of a pre-registered threshold for enrolling enough participants. The decision to expand to 24-hour operation in January 2019, which was approved by both R.P.H. and 911 call centre director K. Holmes, had been made before the decision to extend the trial.

At this point, the call taker's computer-aided dispatch (CAD) system uniformly randomized half of enrollees to a control group and half to treatment. The CAD system then informed the call taker whether they should follow the business-as-usual ambulance request (the control group) or transfer the caller to a nurse with a warm hand-off (the treatment group). To not interfere with time-sensitive 911 operations, protocols and procedures were developed specifically for the accommodation of this randomization scheme. American Medical Response (now a subsidiary of Global Medical Response) was contracted to provide the nurse triage services.

For the treatment group, the nurse then performed a further triage over the phone using Global Medical Response's proprietary nurse triage system to refine the acuity of the complaint, after which they identified appropriate care options, such as visiting a primary care doctor at a future date, immediately visiting an urgent care clinic or immediately being transported via ambulance to an ED[22]. If appropriate, nurses coordinated non-emergent transportation (for example, a taxi ride) to the recommended care site. If the caller requested an ambulance, nurses would transfer the call back to dispatch, and an ambulance would transport the caller to an ED. The programme also included a habit-formation component whereby nurses encouraged the caller to stay connected with their primary care provider and provided instructions on the appropriate use of 911.

Ensuring community comfort and system capacity in the context of the intervention

One of the main goals of the intervention was to reduce patient crowding in EDs and to reduce the overtaxation of EMS resources and providers. At the same time, those patients often still need to visit some medical professional, even if not an emergency physician. The implementing agencies therefore conducted extensive outreach to residents before the NTL was implemented. The agencies also set up infrastructure to ensure that there were sufficient transportation and primary care services to both satisfy caller demand and minimize any cost burden on callers.

Before programme implementation, DC FEMS conducted a Voice of the Consumer Survey in clinics around DC. In this survey, 80% of respondents indicated that they would feel comfortable speaking with a nurse by phone, 78% indicated that they would speak on the phone with a nurse for more than ten minutes, 90% indicated that they would agree to transportation to a non-ED health care facility if the nurse arranged it and 69% indicated a willingness to receive care at a facility other than their usual site of care.

Before the intervention, emergency vehicles that responded to 911 calls (informally referred to as lights-and-sirens) consisted of the city's ambulances, third-party ambulances, non-emergency vehicles armed with medical expertise and Metropolitan Police Department vehicles. In many cases, a first-responding unit is also dispatched as part of a call response.

After the intervention, non-emergent transportation consisting of regular for-hire vehicles was added to this transportation fleet. To guarantee that these rides were covered by public insurance, DC Medicaid provisions for non-emergent transportation were adjusted. We were not able to track the use of non-emergent transportation, but future research could more precisely quantify uptake and cost savings relative to ambulance utilization.

To meet the new primary and urgent care demand expected to be generated by the intervention, 16 FQHCs were identified. Prior to the launch of the programme, these FQHCs updated their operations so that their hours and capacity would be sufficient to handle the anticipated call volume. Accordingly, Medicaid Managed Care Organizations confirmed that payments for services rendered in any of these FQHCs were allowed without restriction.

Finally, the DC Council conducted several public hearings on the law authorizing and funding the intervention. The council's oversight continued throughout the intervention.

Safety analysis

For the first six months of the intervention, 911 call centre staff reviewed the audio of all calls handled by the triage line, and R.P.H. (in his role as medical director of DC FEMS) reviewed the nurses' notes from all calls. After the first six months, R.P.H. continued reviewing all nurses' notes, while 10% of all calls' audio was reviewed by staff, as well as 100% of all calls ending in a specific diagnosis, with the diagnosis rotating each month. These reviews were used for ongoing quality improvement, as well as training and daily feedback for individual nurses. At the conclusion of the study, all calls for which American Medical Response's software had tracked any interaction with the nurse line (n = 3,182) were investigated by DC FEMS and the 911 call centre. Extended Data Fig. 3 shows the results.

Outcome measures

The programme was designed so that the eligible population was typically experiencing conditions best treated in a primary or walk-in urgent care setting. In line with these goals, stakeholder meetings determined the following salient outcomes: (1) reductions in ambulance utilization given the low-acuity nature of the calls, (2) reductions in ED visits for non-emergent conditions or conditions treatable by PCPs, (3) increases in visits to PCPs, (4) reductions in health care expenditures, (5) repeat calls to 911, (6) self-reported patient satisfaction, (7) financial hardship and (8) savings to the overall 911 system. However, data to examine the latter two outcomes were ultimately unavailable due to agency data constraints.

We registered nine primary outcomes based on the first five salient outcomes (NCT05589168): the proportion of calls resulting in (1) an ambulance dispatch or (2) transport; the proportion of calls resulting in an ED visit that was classified as non-emergent or primary care treatable, either within (3) 24 hours or (4) six months of the call; the proportion of calls resulting in a primary care visit within (5) 24 hours or (6) six months of the call; the proportion of callers who called 911 again within (7) 24 hours or (8) six months; and (9) the change in log-transformed Medicaid expenditures. In this paper, we also report on an exploratory survey of self-reported patient satisfaction.

To examine ambulance utilization, we directly used dispatch data from DC's CAD system. These data reflect a time-ordered sequence of events in response to a call.

To examine ED visits, PCP utilization and health care expenditures, we followed refs. 14,16,23 and used Medicaid claims data to track enrollees who were also Medicaid beneficiaries for two time windows subsequent to their randomization: the 24 hours after their call and the six months after their call. This latter outcome includes both the 24-hour window after the call and the remaining days within a six-month period. It was intended to capture longer-run impacts beyond the proximate window following the call; as we note in the Discussion, the interpretation was complicated by the fact that each call at time $t + 1$ could come either from a new caller freshly randomized or from a previous caller, already randomized at time t.

To measure self-reported patient satisfaction, 24 hours after the initial 911 call, nurses attempted to contact patients who had interacted with a nurse at the phone number they initially used to call 911. If the patient did not respond, they tried again on each of the three subsequent days.

Due to the time-sensitive nature of 911 call responses, CAD records did not contain sufficiently clean data to reliably identify repeat 911 callers. We thus cannot report on primary outcomes 7 and 8. This also affects the structure of our pre-registered statistical analysis (see 'Statistical analyses').

While our primary analyses examined overall effectiveness across the entire duration of the intervention, we provide the results of regressions that estimate month-specific effects in Extended Data Table 1. We pre-registered this month-by-month analysis to account for the fact that the intervention's impacts might change over time—for instance, as callers establish connections with PCPs or as the nurse triage programme model matures. Ultimately, we found similar impacts across each month of the study.

Political and executive stakeholders, along with all partner agencies' directors, reviewed and signed off on the publication of the pre-analysis plan, and also agreed to pre-register discussions on the interpretation of potential results (see the 'Policy implications' section of the pre-analysis plan). Below, we discuss the details on how each of the outcome variables are defined.

Identifying the treatment and control groups. Whether a caller was eligible and enrolled in the trial was recorded in DC's CAD system. Between 19 April 2018 and 1 March 2019, the system recorded that 8,761 callers were eligible on the basis of their chief complaint. However, 2,708 were excluded from the study, primarily because they were calling outside of operating hours (n = 2,033) or because the call was cancelled after the chief complaint was determined (n = 489). In total, 6,053 callers entered the randomized evaluation of DC's 911 nurse triage programme. Of these, 3,030 were assigned to the nurse (the treatment group) and 3,023 were assigned to business-as-usual (the control group). Figure 1 shows the participant flow.

We note that this number of treated callers (3,030) differs from the number of callers followed in our safety analysis (3,182). As best as we can determine, this is due to a syncing issue between the software that triage nurses use and DC's CAD system. While the CAD system is the system of record for the randomization, for the safety analysis, we investigated any call for which the nurse triage system had a record, a larger group than our analytic sample.

As a check that our identification strategy was correct, Extended Data Fig. 1 shows the monthly aggregate counts of treatment and control group callers. These are well matched over time, with closer matches as the call volume increases. The figure shows that calls referred to the NTL increased over time, with a particular increase after the programme expanded to 24-hour-a-day operation in January 2019.

Once the treatment and control groups had been established, we had to further identify whether an ambulance dispatch had resulted in a transport.

Defining ambulance utilization. Ambulance utilization could refer to two potential uses of ambulances: whether the ambulance was dispatched in response to a call or whether the dispatched ambulance also transported a caller. The latter typically represents a much larger investment of resources. We report the results for both operationalizations.

DC's CAD system records a sequence of time-stamped events that reflect what happened as paramedics searched for and ultimately provided assistance to the caller. For our primary results, we operationalized 'ambulance transport' as occurring when the final event in this sequence indicated that a transport had occurred. However, sometimes a transport was notated but was not the final event. As indicated in Table 1, our results remained robust if we instead operationalized using any event indicating a transport.

Identifying Medicaid beneficiaries. Our next outcomes looked at the subpopulation of callers who were Medicaid beneficiaries. To determine which callers were Medicaid beneficiaries, DC's Department of Health Care Finance (DHCF) performed 'fuzzy matching' between a caller's recorded first and last names and their date of birth and all Medicaid enrolment information. This matching requires an exact match on date of birth and uses the SOUNDEX function to look for highly similar-sounding first and last names between the caller and the Medicaid beneficiary.

Clearly, this matching strategy requires the presence of both the caller's first and last name and their date of birth. Of the 6,053 callers in our study, 1,012 were missing at least one of these identifiers. Of the remaining 5,041 callers, 3,648 matched to a Medicaid record. Those matched to a Medicaid record are not a random sample of callers. Instead, they may represent a 'middle range' of callers who are able to navigate administrative processes to qualify for Medicaid, thus representing a more advantaged subset than uninsured callers, but who might face worse underlying health statuses than callers with employer-sponsored insurance. Supplementary Information section 2 discusses (1) the match rates between the treatment and control callers and (2) the demographic characteristics of these beneficiaries.

Defining ED utilization. To determine whether a Medicaid beneficiary used an ED and whether that visit could have been avoided, we classified the ICD-10 codes appearing in their claim history using the NYU Center for Health and Public Service Research ED utilization coding scheme[24]. The coding scheme classifies codes into one of four buckets:

1. Non-emergent
2. Emergent/primary care treatable
3. Emergent–ED care needed–preventable/avoidable
4. Emergent–ED care needed–not preventable/avoidable

As primary outcomes, we report whether a caller used an ED within 24 hours and within six months of their call, and the coding scheme classified their diagnosis as one of 'non-emergent' or 'emergent/primary care treatable'.

Defining primary care utilization. To determine whether a Medicaid beneficiary used a PCP, we classified the ICD-10 codes appearing in their claim history using a classification scheme provided by DC's DHCF, the district's Medicaid office. We note that our pre-analysis plan stated that this classification would "rely on the same algorithm as with Emergency Department use". However, since the NYU classifications we used for ED use do not cover primary care, we used DHCF's classification scheme.

Similar to the ED visits outcome, we report whether a caller used primary care services within 24 hours and six months of their call.

Defining health care expenditures. We also examined health care expenditures for each beneficiary in the 24-hour and six-month windows after their call. We used a classification scheme developed by DC's DHCF that involves filtering out payments made as part of capitated payments/managed care and avoiding double-counting expenditures by separating claims processed at the header level from claims processed at the non-header, detail level, and then recombining these sums to get a beneficiary-level aggregate.

Power analysis

To perform a power analysis, we first gathered historical data on 911 calls that would probably have been eligible for the NTL on the basis of their chief complaints (see Appendix G for our pre-analysis plan). Because DC was simultaneously migrating to a new system of chief complaints, this process required experts at DC FEMS to match chief complaints in the old system to chief complaints in the new system[13]. We described this as 'fuzzy matching' in our pre-analysis plan. However, this fuzzy matching was distinct from the matching needed to link callers to Medicaid beneficiary files. In particular, the matching

of old complaints to a new set of chief complaints was done manually, and the 'fuzzy' refers to the fact that there is not exact correspondence between the old and new systems. This allowed us to compute baseline outcomes for these callers as well as estimate how many calls we might expect to be enrolled (see especially Appendix C of the pre-analysis plan estimating nurse supply elasticity). We then computed minimal detectable effect sizes on the basis of a simple two-tailed t-test for continuous outcomes and a two-tailed z-test for binary outcomes, setting $\alpha = 0.05$ and $\beta = 0.80$, and making a Bonferroni correction for eight hypothesis tests. On the basis of conversations with stakeholders, we determined that a sample size of 5,400 participants would allow us to detect meaningful effects. On the basis of our analysis of historical data, we believed that we would easily surpass this number in a six-month trial, though if we hit our worst-case projections (approximately 20 calls per day), we committed to extending the trial by three months to hit 5,400 participants. Unfortunately, after six months, the number of eligible calls was below our worst-case estimates due to dispatchers needing time to become more comfortable using the system. R.P.H., in consultation with stakeholders across the DC government, subsequently decided to extend the experiment timeline to 1 March 2019, or approximately 10.5 months after launch. During this period, data analysts remained blinded to the outcome data, though they were aware of the total number of eligible callers in the control and treatment groups.

Statistical analyses

Our pre-analysis plan proposed a regression analysis that accounted for (a) repeat callers, (b) pretreatment covariates of callers and (c) the possibility of different treatment assignment probabilities by day that might result from operational considerations (for example, a staff member departure that took longer than expected to fill). Ultimately, we could not, from the CAD data alone, reliably match a call to an individual, and thus we could not reliably determine whether someone was a repeat caller or their pretreatment covariates. Moreover, the situation contemplated in possibility (c) never occurred. Our pre-specified analysis thus collapses to:

$$y_c \sim t_c$$

where y_c is the observed outcome for call c and t_c is the treatment indicator for call c. We note that this is a departure from our plan in that we are now examining results at the call and not the individual level. This also precludes us from reporting any dosage effects.

Three of our four non-satisfaction outcomes y_c are binary, so we report the results of a logistic regression of y_c on t_c, including the associated point estimate and the 95% confidence interval from this regression. For the one continuous outcome of health care expenditures, we report the results of a two-tailed linear regression. Due to the right-skewed nature of expenditures (many beneficiaries with US$0 in expenditures and a few with high expenditure values), we defined the outcome measure as the log-transformed total expenditures.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The data analysed in this paper were provided by DC's FEMS, Office of the Chief Technology Officer and DHCF and contain protected health information. To protect privacy, we cannot publicly post individual-level data. Qualified researchers and relevant approvals including ethical approval can request access to the de-identified data about this trial from the corresponding author. A formal contract will be signed, and an independent data protection agency should oversee the sharing process to ensure the safety of the data.

Code availability

All code used to produce this analysis is publicly available at https://github.com/thelabdc/FEMS-911NurseTriageLine-public.

References

1. *9-1-1 Statistics* (National Emergency Number Association, 2020).
2. Sporer, K. A. 911 patient redirection. *Prehosp. Disaster Med.* **32**, 589–592 (2017).
3. *Reducing Emergency Department Overuse: A $38 Billion Opportunity* (New England Health Institute, 2011).
4. Bruijns, S. R. et al. Effect of introduction of nurse triage on waiting times in a South African emergency department. *Emerg. Med. J.* **25**, 395–397 (2008).
5. Mills, J. et al. Effectiveness of nurse triage in the emergency department of an urban county hospital. *J. Am. Coll. Emerg. Physicians* **5**, 877–882 (1976).
6. *A Model for Better Community Healthcare: How One EMS System Achieved the Triple Aim from a Federal Health Care Innovation Award Grant* (REMSA, 2017).
7. Gardett, I. et al. 911 emergency communication nurse triage reduces EMS patient costs and directs patients to high-satisfaction alternative point of care. *Ann. Emerg. Dispatch Response* **3**, 8–13 (2015).
8. Møller, T. P. et al. Why and when citizens call for emergency help: an observational study of 211,193 medical emergency calls. *Scand. J. Trauma Resusc. Emerg. Med.* **23**, 88 (2015).
9. Zinger, N. D., Blomberg, S. N., Lippert, F., Krafft, T. & Christensen, H. C. Impact of integrating out-of-hours services into Emergency Medical Services Copenhagen: a descriptive study of transformational years. *Int. J. Emerg. Med.* **15**, 40 (2022).
10. Eastwood, K. et al. A novel approach for managing the growing demand for ambulance services by low-acuity patients. *Aust. Health Rev.* **40**, 378–384 (2015).
11. *District of Columbia Fire and Emergency Medical Services Integrated Healthcare Collaborative, Final Report* (District of Columbia Fire & Emergency Medical Services, 2017).
12. Berchick, E. R. et al. *Health Insurance Coverage in the United States: 2018* (US Department of Commerce, 2019).
13. DeLeire, T. in *Health and Labor Markets* (eds. Tatsiramos, K. & Polachek, S. W.) 155–194 (Emerald, 2019).
14. Finkelstein, A. N. et al. Effect of Medicaid coverage on ED use—further evidence from Oregon's experiment. *N. Engl. J. Med.* **375**, 1505–1507 (2016).
15. Lavetti, K. J. et al. *How Do Low-Income Enrollees in the Affordable Care Act Marketplaces Respond to Cost-Sharing?* Technical Report No. 26430 (National Bureau of Economic Research, 2019).
16. Taubman, S. L. et al. Medicaid increases emergency-department use: evidence from Oregon's health insurance experiment. *Science* **343**, 263–268 (2014).
17. Datta, S. & Mullainathan, S. Behavioral design: a new approach to development policy. *Rev. Income Wealth* **60**, 7–35 (2014).
18. Rose, A. J. et al. Primary care visit regularity and patient outcomes: an observational study. *J. Gen. Intern. Med.* **34**, 82–89 (2019).
19. Daniels, O. District triage line continues to trim non-emergency calls from 911. *Washington Post* 20 November (2023).
20. Python Language Reference, v.3.8.6 http://www.python.org (Python Software Foundation, 2020).
21. R: a language and environment for statistical computing https://www.R-project.org/ (R Foundation for Statistical Computing, 2021).
22. Briggs, J. K. *Telephone Triage Protocols for Nurses* (Wolters Kluwer Health, 2020).
23. Finkelstein, A. et al. Health care hotspotting—a randomized, controlled trial. *N. Engl. J. Med.* **382**, 152–162 (2020).

24. Billings, J. et al. *Emergency Room Use: The New York Story* (Commonwealth Fund, 2000).

Acknowledgements

This work would not be possible without the support of M. Bowser (mayor), K. Donahue (city administrator), R. Young (former city administrator), S. Quinney (director of The Lab @ DC), G. M. Dean (retired fire chief), FEMS, the Office of Unified Communications, the DHCF, the FQHCs, the DC Primary Care Association, the Department of Health and the Office of the Deputy Mayor for Public Safety and Justice. We thank the following for their time and assistance in making this evaluation possible: M. Baisley, A. Beaton, V. Bishop, D. Braman, R. Breslin, S. Brown, I. Bucksell, J. Coombs, D. Corcoran, T. Curtis, O. Dedner, J. Doleac, N. Donnelly, J. Duff, T. Dutta, L. Edmonson, B. Egar, K. Gan, H. Gil, A. Grady, J. Greenberg, K. Holmes, E. Holve, A.-T. Huang, A. Huberts, T. Kavaleri, D. Kornfield, E. Koshkin, B. Kreiswirth, B. Krucoff, K. Liebowitz, M. MacCarthy, K. Minnich, A. Mauro, L. Nesbitt, C. Nguyen, C. Nwaete, P. Oandasan, K.Y. Oh, J. Reed, M. Reed, S. Roque, C. Scholsberg, N. Smith, D. Stanescu, P. Testa, T. Thangalvadi, W. Turnage, J. Wedeles, D. Weinroth, J. Weissfeld, O. Whittaker, M. Williamson, S. Willig and J. Wobbleton. We especially thank M. A. Bates, N. Choudhry and B. Özler for reviewing the pre-analysis plan and M. Welch for her substantive and project management contributions during the review process. The intervention as a whole was funded by general appropriations from the District of Columbia. R.P.H., as the medical director of DC FEMS, was also funded through DC appropriations. C.H., K.H.W., R.T.M. and D.Y. were funded through the Laura and John Arnold Foundation (now Arnold Ventures). R.A.J. was funded through the ABF/JPB Foundation Access to Justice Scholars programme for time supporting the work. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

Conceptualization: C.H., R.P.H., R.T.M., K.H.W. and D.Y. Methodology: C.H., R.T.M., K.H.W. and D.Y. Visualization: R.A.J. and K.H.W. Coding and analysis—original analysis: R.A.J. and K.H.W. Coding and analysis—review and supplemental: R.A.J., R.T.M. and K.H.W. Project administration: C.H. and R.P.H. Supervision: C.H. and R.T.M. Writing—original draft: C.H., R.A.J. and K.H.W. Writing—review and editing: C.H., R.P.H., R.A.J., R.T.M., K.H.W. and D.Y.

Competing interests

R.P.H. was during the commission of this study the medical director of DC FEMS and draws a salary from the District of Columbia. The remaining authors declare no competing interests.

Additional information

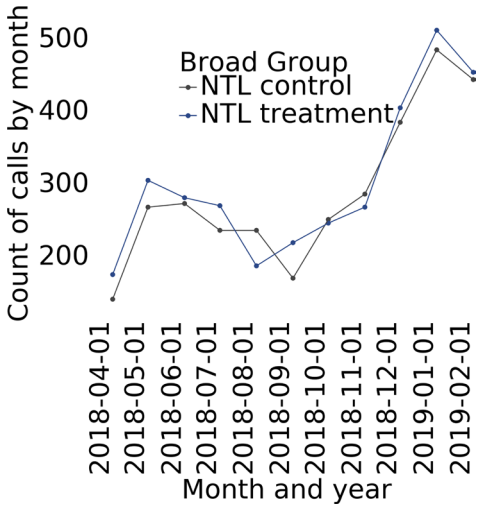
Extended data is available for this paper at https://doi.org/10.1038/s41562-024-01889-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-024-01889-6.

Correspondence and requests for materials should be addressed to Kevin H. Wilson.

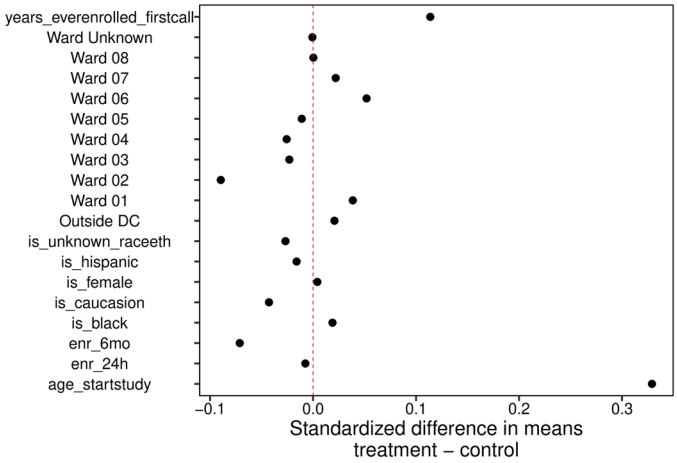
Peer review information *Nature Human Behaviour* thanks Stefan Morreel and Armann Ingolfsson for their contribution to the peer review of this work.

https://doi.org/10.1038/s41562-024-01889-6



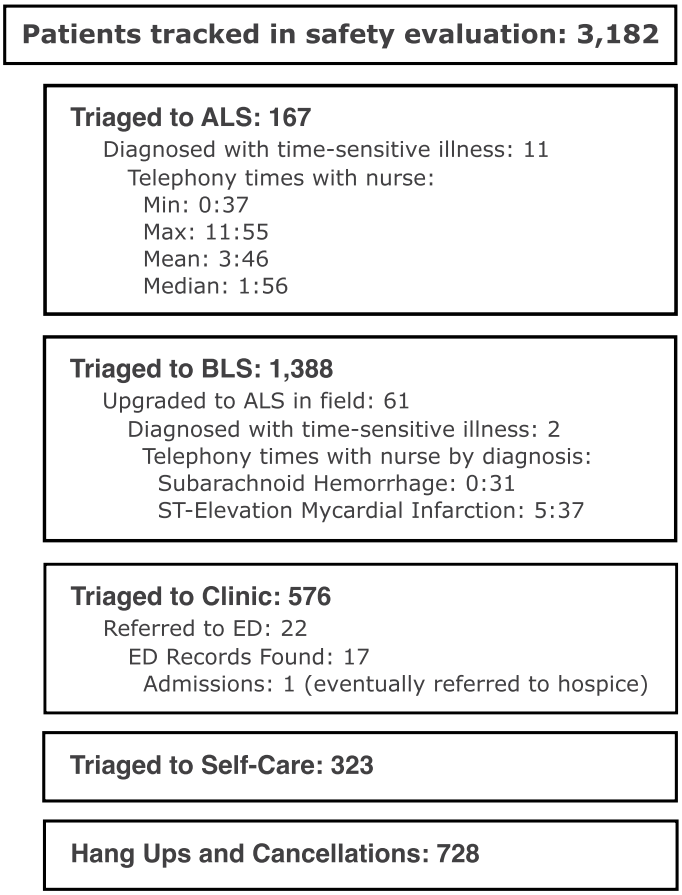
Extended Data Fig. 1 | Overall counts of treated and control group callers aggregated to the monthly level. The figure shows the raw counts of treatment and control group callers for each of the eleven months in the study period. We see a sharp rise after the triage line was expanded to 24 hours per day.

https://doi.org/10.1038/s41562-024-01889-6



Extended Data Fig. 2 | Balance plot: treatment and control group (Medicaid beneficiary sample). The figure shows the standardized mean difference for each attribute between the treatment and control group, which helps us compare variables on different scales (for example, years versus percentages).

https://doi.org/10.1038/s41562-024-01889-6



Extended Data Fig. 3 | Detailed outcomes of the safety analysis. The flow chart shows a low incidence of safety events.

https://doi.org/10.1038/s41562-024-01889-6

Extended Data Table 1 | Treatment effects on ambulance use by month

Month	Treatment effect	p-value (uncorrected)	p-value (corrected)
2018-04	-0.8423	0.0006407084	0.0070477927
2018-05	-0.9477	0.0000003074	0.0000033811
2018-06	-1.4311	0.0000000000	0.0000000000
2018-07	-0.9430	0.0000008239	0.0000090630
2018-08	-1.9095	0.0000000000	0.0000000000
2018-09	-1.2656	0.0000000221	0.0000002431
2018-10	-1.3972	0.0000000000	0.0000000001
2018-11	-1.3573	0.0000000000	0.0000000000
2018-12	-1.4249	0.0000000000	0.0000000000
2019-01	-1.0990	0.0000000000	0.0000000000
2019-02	-0.9907	0.0000000000	0.0000000001

The treatment effect column of each row shows the coefficient from a logistic regression that regresses the binary measure of ambulance use on the treatment indicator. P-values are from a two-tailed test of the null hypothesis that the treatment coefficient is equal to zero. The uncorrected p values are the original p values from the regression. The corrected p values apply a Bonferroni adjustment, implemented with this function (https://www.rdocumentation.org/packages/stats/versions/3.6.2/topics/p.adjust) and with the number of comparisons equal to the number of months (11). The results show consistent, statistically significant effects in reducing ambulance use.

https://doi.org/10.1038/s41562-024-01889-6

Extended Data Table 2 | Breakdown of identifiers by ambulance status

	Dispatch and transport	Dispatch but no transport	No dispatch or transport
Control group callers			
Has both DOB and name	0.97	0.63	0.01
Missing dob; has name	0.03	0.02	0.00
Missing dob and name	0.00	0.35	0.99
Treatment group callers			
Has both DOB and name	0.99	0.79	0.61
Missing dob; has name	0.01	0.03	0.03
Missing dob and name	0.00	0.18	0.36

The table shows, separately for the treatment and control groups, the presence of identifiers by three types of dispositions: an ambulance was dispatched and transported the caller; an ambulance was dispatched but there was no transport; and an ambulance was not dispatched. We see that both groups have similar coverage rates for names and DOBs once we take into account the relationship between ambulance utilization and these coverage rates.

https://doi.org/10.1038/s41562-024-01889-6

Extended Data Table 3 | Treatment and control group characteristics: Medicaid sample

Attribute	Treatment (mean or proportion)	Control (mean or proportion)	Difference + 95% CI
Age	48.00	45.72	2.28 [-3.7972, -0.7629]
Medicaid enrollment (days w/in 24 hours post call)	0.97	0.98	-0.0074 [-0.0043, 0.0191]
Medicaid enrollment (days w/in 6 mo. post call)	176.59	177.54	-0.9461 [-0.9749, 2.867]
Race/eth: Black	0.86	0.85	0.0175 [-0.0439, 0.0089]
Race/eth: Non-hispanic white	0.02	0.02	-0.0056 [-0.0048, 0.0161]
Female	0.59	0.58	0.0032 [-0.0401, 0.0337]
Race/eth: Hispanic	0.02	0.02	-0.0021 [-0.0081, 0.0124]
Race/eth: Unknown	0.09	0.10	-0.008 [-0.0138, 0.0299]
Outside DC	0.00	0.00	0.001 [-0.0043, 0.0022]
Ward 01	0.07	0.06	0.0105 [-0.0295, 0.0085]
Ward 02	0.09	0.11	-0.0266 [0.0041, 0.049]
Ward 03	0.01	0.01	-0.0021 [-0.0052, 0.0095]
Ward 04	0.08	0.08	-0.0071 [-0.0133, 0.0276]
Ward 05	0.13	0.13	-0.0039 [-0.0213, 0.0292]
Ward 06	0.12	0.11	0.0183 [-0.0422, 0.0056]
Ward 07	0.21	0.20	0.01 [-0.0402, 0.0201]
Ward 08	0.22	0.22	0.0001 [-0.0313, 0.0311]
Ward Unknown	0.07	0.07	-0.0002 [-0.0183, 0.0187]
Medicaid enrollment (years before call)	15.57	15.12	0.4492 [-1.0969, 0.1985]

The table shows the characteristics of the treatment and control group callers. The last column shows the difference in values accompanied by the 95% confidence interval for a two-tailed test of differences in the means or proportions. We see that for nearly all the characteristics, the confidence interval crosses zero, indicating we fail to reject the null of no difference in characteristics between the treatment and control group callers.

https://doi.org/10.1038/s41562-024-01889-6

Extended Data Table 4 | Exact rates and proportions: Emergency department and primary care physician visits

Outcome	Time horizon	Disposition	Percent
ED visit (non-emergent; 24 hours)	24 hours	NTL control	29.70
ED visit (non-emergent; 24 hours)	24 hours	NTL treatment	25.14
ED visit (non-emergent; 6 months)	6 months	NTL control	43.46
ED visit (non-emergent; 6 months)	6 months	NTL treatment	43.26
Primary care visit (24 hours)	24 hours	NTL control	2.50
Primary care visit (24 hours)	24 hours	NTL treatment	8.22
Primary care visit (6 months)	6 months	NTL control	43.34
Primary care visit (6 months)	6 months	NTL treatment	42.77

The table shows the exact rates of emergency department (ED) and primary care physician visits across the different treatment and control groups and time horizons.

nature portfolio

Corresponding author(s): Rebecca A. Johnson

Last updated by author(s): Feb 29, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒	A description of all covariates tested
☐	☒	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Data was collected through DC's computer-aided dispatch system, its ePCR system (SafetyPAD), its Medicaid claims system, and American Medical Response's nurse triaging system (Logis). We do not have specific version numbers of the software used by these government agencies.
Data analysis	Data was analyzed utilizing Python (version 3.8.6) and R (version 4.1.2). All packages are open source and freely available. Exact packages and their versions may be found in the poetry.lock file (Python) and renv.lock (R) file of our pubilc code repository, available here: https://github.com/thelabdc/FEMS-911NurseTriageLine-public

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

The data analysed in this paper were provided by DC's Fire and Emergency Medical Services, Office of the Chief Technology Officer, and Department of Health Care Finance and contains protected health information. To protect privacy, we cannot publicly post individual-level data. Qualified researchers with a valuable research question and relevant approvals including ethical approval can request access to the de-identified data about this trial from the corresponding author. A formal contract will be signed and an independent data protection agency should oversee the sharing process to ensure the safety of the data.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	Due to the emergency posture and call-level nature of our intervention, sex, gender, race, age, and other demographic information was not able to be collected for our overall study population. However, we have provided demographic breakdowns along several dimensions for our Medicaid sample in Extended Data Figure 7. These data are administrative in nature, and so reflect self-reported race and gender information either from the time of Medicaid (re-)enrollment or potentially from a correction action initiated by either Medicaid or the beneficiary. In particular, they do not necessarily reflect the most up-to-date information about a beneficiary's race or gender at the time of their enrollment into the study.
Population characteristics	See above.
Recruitment	When a 911 call was placed, a call taker would perform an initial assessment of the caller's chief complaint utilizing the Criteria-Based Dispatch system developed by King County, Washington. If the caller was a non-incarcerated adult calling on behalf of themselves (a "first-party caller"), the complaint was eligible to be serviced by the triage line, and there was a nurse available, then the caller was enrolled in our study. At this point, the call taker's computer-aided dispatch (CAD) system uniformly randomized half of enrollees to a control group and half to treatment. The CAD then informed the call taker whether they should follow the business-as-usual ambulance request (control group) or transfer the caller to a nurse with a warm hand off (treatment group). Since the procedure was applied to the full universe of eligible 911 calls, we do not anticipate any selection biases in which calls/callers ended up in the sample. See Methods Section 4.1 for more details.
Ethics oversight	The study protocol was reviewed by the Western Institutional Review Board, which confirmed it adhered to standards of ethical research in WIRB Work Order 1-1033362-1. This notation also appears as the first sentence of the Methods section.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☐ Life sciences ☒ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	Nurses triaged first-party calls from unincarcerated adults who presented one of several low-acuity chief complaints. Nurses could provide non-emergent transportation to a Federally Qualified Health Center, recommend self-care, or return callers to the traditional 911 system. Between April 19, 2018, and March 1, 2019, 6,053 callers were randomized (1:1) to receive a business-as-usual response or further triage from the nurse.
Research sample	The research population was all callers to DC's 911 system during the relevant time period who were eligible to be enrolled. All eligible callers were enrolled in the study. Demographics were collected for the subset of callers who were matchable to DC's Medicaid beneficiary file. Among this subset,and focusing on the control group, the mean age was 45.7 years old, 85% of the callers self-identified as Black, and 58% of the callers self-identified as female.
Sampling strategy	All callers who were eligible were enrolled in the study. Randomization was performed in real time by the District's CAD system. A

Sampling strategy	detailed power analysis, including justifications for our desired population size, is provided in Section 4.5 of the supplementary Methods text.
Data collection	Data was collected through DC's computer-aided dispatch system, its ePCR system (SafetyPAD), its Medicaid claims system, and American Medical Response's nurse triaging system (Logis). CAD data represents the real time actions of emergency personnel, including call takers and paramedics. The ePCR system represents field notes of emergency personnel. Logis represents nurses' notes and actions throughout the trial. And the Medicaid claims system represents requests for reimbursements by medical staff and emergency personnel, usually filed well after the actual intervention. None of the researchers were present for the data collection and the employees entering data into the administrative systems used for data collection were blinded to the participant's experimental condition.
Timing	The trial ran from April 19, 2018, to March 1, 2019,
Data exclusions	No data were excluded.
Non-participation	No participants dropped out.
Randomization	Randomization was Bernoulli and performed after a call taker determined a caller was eligible to be enrolled in the nurse triage program.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☐	☒ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Clinical data

Policy information about clinical studies

All manuscripts should comply with the ICMJE guidelines for publication of clinical research and a completed CONSORT checklist must be included with all submissions.

Clinical trial registration	NCT05589168
Study protocol	https://osf.io/xderw
Data collection	The data was collected from DC administrative data systems (the computer-aided dispatch system and the Medicaid beneficiary files/ claims data) upon the trial's completion in March of 2019. These data systems capture all calls dispatched through DC's 911 system and all health services billed to DC Medicaid agency (regardless of the site of the healthcare provision).
Outcomes	We report on seven pre-registered primary outcomes: changes in the rate of (1) ambulance dispatches and (2) transports from 911 calls; changes in the rates of Medicaid beneficiaries' visits to primary care physicians within (3) 24 hours and (4) 6 months of interacting with the nurse help line; changes in the rate at which Medicaid beneficiaries subsequently receive a diagnosis within (5) 24 hours and (6) 6 months from an emergency department where the diagnosis's ICD-10 code indicates it was likely "non-emergent" or "primary-care treatable;" and (7) total log-transformed Medicaid expenditures within six months of a 911 call. Two of our pre-registered primary outcomes—measuring the rate of repeat callers to the 911 system between the control and treatment groups over two different time periods—were ultimately infeasible due to insufficient data collected at the time of the call.