

RESEARCH

Open Access



# Understanding EMS response times: a machine learning-based analysis

Peter Hill<sup>1,4\*</sup>, Jakob Lederman<sup>1,4</sup>, Daniel Jonsson<sup>5</sup>, Peter Bolin<sup>1,6</sup> and Veronica Vicente<sup>2,3,4</sup>

## Abstract

**Background** Emergency Medical Services (EMS) response times are critical for optimizing patient outcomes, particularly in time-sensitive emergencies. This study explores the multifaceted determinants of EMS response times, leveraging machine learning (ML) techniques to identify key factors such as urgency levels, environmental conditions, and geographic variables. The findings aim to inform strategies for enhancing resource allocation and operational efficiency in EMS systems.

**Methods** A retrospective analysis was conducted using over one million EMS missions from Stockholm, Sweden, between 2017 and 2022. Advanced ML techniques, including Gradient Boosting models, were applied to evaluate the influence of diverse variables such as call handling times, travel times, weather patterns, and resource availability. Feature engineering was employed to extract meaningful insights, and statistical models were used to validate the relationships between key predictors and response times.

**Results** The study revealed a complex interplay of factors influencing EMS response times, aligning with the study's aim to deepen the understanding of these determinants. Key drivers of response time variability included weather conditions, call priority, and resource constraints. ML models, particularly Gradient Boosting, proved effective in quantifying these impacts and provided robust predictions of response times across scenarios. By providing a comprehensive evaluation of these influences, the results support the development of adaptive resource allocation models and evidence-based policies aimed at enhancing EMS efficiency and equity across all call priorities.

**Conclusions** This study underscores the potential of ML-driven insights to revolutionize EMS resource allocation strategies. By integrating real-time data on weather, call types, and workload, EMS systems can transition to adaptive deployment models, reducing response times and enhancing equity across priority levels. The research provides a blueprint for implementing predictive analytics in EMS operations, paving the way for evidence-based policies that improve emergency care efficiency and outcomes.

**Clinical trial number** Not applicable.

**Keywords** Emergency medical services, Response times, Machine learning, Resource allocation, Predictive analytics, Emergency care optimization

\*Correspondence:

Peter Hill

peter.hill@ki.se

<sup>1</sup>Region Stockholm Health and Medical Care Administration, The Department for Specialized Care, Stockholm, Sweden

<sup>2</sup>The Ambulance Medical Service in Stockholm (AISAB), Stockholm, Sweden

<sup>3</sup>Academic EMS in Stockholm, Stockholm, Sweden

<sup>4</sup>Karolinska Institutet, Department of Clinical Science and Education, Södersjukhuset in Stockholm, Stockholm SE-118 83, Sweden

<sup>5</sup>KTH Royal Institute of Technology, The Department of Urban Planning and Environment, Stockholm, Sweden

<sup>6</sup>Karolinska Institutet, Department of Learning, Informatics, Management and Ethics (LIME), Solna, Sweden



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/>.

## Background

Emergency medical services (EMS), integral to out-of-hospital emergency care, utilize specialized vehicles and exhibit a notable degree of organizational similarity across different countries [1, 2]. The EMS process begins when an emergency call is received by the Emergency Medical Communication Centre (EMCC), where an Emergency Medical Dispatcher (EMD) assigns a priority based on the perceived risk of the patient having a time-critical medical condition [1, 3]. In Sweden, there are three priority levels: Priority 1, indicating the highest urgency, followed by Priorities 2 and 3 for lower levels, with lower-priority cases being deferred if resources are limited [3].

Accurately assessing the urgency of emergency calls is challenging due to factors such as language barriers and caller agitation. Emergency calls are rarely made by patients themselves and are often made by relatives or bystanders, leading to discrepancies between the urgency assigned by the EMCC and the clinical assessment performed by EMS personnel upon arrival. These discrepancies can result in misjudgment, potentially impacting patient outcomes [4–8].

Response time, defined as the interval from the initiation of an emergency call to the arrival of EMS personnel on the scene, is a globally recognized metric for evaluating EMS performance, particularly in high-urgency cases where lights and sirens are used. It comprises two phases: call handling time, from receipt of the call to the dispatch of resources, and travel time, from resource deployment to arrival at the scene [9]. Delays in response time can affect patient outcomes, making it a critical factor in EMS resource planning. Resource shortages in specific geographic areas often exacerbate response times [9–12]. In Sweden, the Swedish Association of Local Authorities and Regions (SKR) employs the median response time for Priority 1 calls as a benchmark for comparative assessment across EMS systems nationwide [13].

Despite extensive research on EMS resource allocation and optimization models, gaps remain in understanding how various factors interact to influence response times. Real-time traffic data, historical trends, and dynamic resource allocation strategies, such as “fluid deployment,” have shown potential for reducing response times. However, many studies have focused on isolated factors like traffic or weather without accounting for their combined effects or the disparities between urgency levels [6, 14–29].

While using lights and sirens accelerates response and transportation times, it poses risks of vehicular accidents and threats to public safety. Evidence supporting response times direct positive impact on patient outcomes remains limited [30]. Additionally, factors such as weather conditions and temporal variations, can impact

response times, patient volumes, and resource deployment, further complicating resource allocation strategies [31–42].

## Advances in machine learning in EMS systems and research gap

Recent advancements in data collection and machine learning (ML) have significantly transformed Emergency Medical Services (EMS) operations. ML technologies have enhanced operational efficiency, minimized response times, and improved patient outcomes. Contemporary research underscores the critical roles of ML in EMS systems, particularly in demand prediction, resource optimization, and the integration of environmental and temporal factors into planning frameworks.

Machine learning models have demonstrated remarkable capabilities in several domains. They effectively forecast EMS demand by leveraging spatio-temporal data, traffic patterns, and even insights from social media [21, 26, 43]. Bayesian models, dynamic relocation strategies, and robust optimization algorithms have improved resource allocation, directly reducing response times [23, 44–46]. Additionally, regression-based models, including gradient boosting techniques, estimate response times with high accuracy, factoring in external variables such as weather and traffic conditions [6, 10, 39].

However, despite these strides, significant challenges and research gaps remain. Most studies have focused on individual factors like traffic or weather, neglecting their combined effects on EMS efficiency [6, 10, 21, 25, 27, 39]. Moreover, the scalability of ML models across different geographic and operational contexts is underexplored [23, 44–47]. Practical implementation of real-time ML predictions in live EMS operations also poses logistical and computational hurdles, especially during high-risk situations [6, 10, 39].

Addressing these gaps, the current study introduces an integrative ML framework that combines various factors such as urgency levels, geographic contexts, and weather conditions to provide a holistic analysis of EMS response dynamics. Unlike prior research, this approach not only evaluates the interplay of these variables but also proposes equitable strategies for resource allocation across different priority levels, emphasizing both high-priority and lower-priority cases.

## Outline of the paper

This paper begins by introducing the current state of knowledge regarding EMS response times, emphasizing both their significance in clinical outcomes and the limitations of solely using traditional metrics such as high-priority response times. It then explains how machine learning techniques can integrate various factors—urgent call assessments, dynamic resource deployment,

and environmental variables—to develop a more holistic understanding of EMS operations. Subsequently, the data sources and methodological framework are discussed, highlighting the retrospective observational design, over one million EMS missions included from Stockholm between 2017 and 2022, the use of feature engineering, and the application of both linear regression and advanced machine learning methods (particularly gradient boosting). The analysis then focuses on the study's main findings, demonstrating how urgency levels, weather conditions, and geographic factors collectively influence EMS response times, while also comparing model performance across different call priorities. This leads into a broader discussion that contextualizes these results within existing literature, explores their potential implications for resource allocation and policy, and addresses the feasibility and challenges of real-time machine learning implementation in EMS. Lastly, the paper concludes by underscoring key contributions, proposing strategies for more equitable and efficient EMS management, and outlining directions for further research, including the integration of missing data sources such as real-time traffic information, refined model deployment, and applicability to other regions.

### Aim

This study aims to deepen the understanding of the complex factors influencing EMS response times. By examining the interplay between urgency levels, environmental conditions, and geographical contexts, it seeks to uncover actionable insights that can optimize resource allocation and enhance the effectiveness of emergency response operations.

## Materials and methods

### Study design

This study utilizes a retrospective observational design, leveraging an extensive dataset of over one million emergency medical service (EMS) missions conducted in Stockholm, Sweden, from 2017 to 2022. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) checklist, designed to assist authors in presenting observational studies with clarity and rigor, was followed to ensure comprehensive and transparent reporting [48]. The invaluable EMS procedural data were sourced from the Region Stockholm VAL databases, which are renowned for their robust and cohesive data at the individual level [49, 50]. Furthermore, to enrich our analysis, meteorological data from the Swedish Meteorological and Hydrological Institute were seamlessly integrated to provide invaluable contextual insights [7, 50, 51]. For analytical rigor, the study employed diverse statistical methods, including linear regression and advanced machine learning models, supported by feature

engineering to incorporate key variables influencing EMS response time. A directed acyclic graph (DAG) was constructed to visually represent factors partially assumed to be influencing response time, elucidating potential causal pathways. To enhance the understanding of the multifaceted factors influencing EMS response time and their interrelationships, data analyses were conducted via a diverse array of statistical methodologies, including both traditional linear regression and advanced machine learning models [52–54]. These analyses were meticulously synthesised, presenting a comprehensive blend of inferential and descriptive statistics. To increase the analytical depth, sophisticated data mining techniques and feature engineering have been employed to identify and incorporate variables known to significantly influence response time [8, 9, 14, 31–37, 39, 40, 42, 55]. The program code used for feature engineering and machine learning is provided in Appendix B.

### Setting of the study

The study was conducted in Stockholm, a metropolitan area with a population of approximately 2.5 million inhabitants [56]. Healthcare services in Stockholm, including EMSs, are overseen by Region Stockholm [56]. The EMCC is operated by SOS Alarm AB, a publicly owned entity. EMS operations are carried out by a combination of private and publicly owned operators contracted by Region Stockholm [56, 57]. The EMS resources are stationed at fixed stations with the option of “fluid deployed” [9, 58–60].

### Participants

The initial dataset comprises a total of  $N=1\,297\,858$  observations, encompassing all EMS missions conducted in Stockholm between 2017 and 2022. Employing the exclusion criteria outlined in Table 1, the dataset underwent a meticulous refinement process, resulting in a refined subset of  $n=1\,144\,754$  observations deemed appropriate for subsequent feature engineering and in-depth analysis.

### Variables and features

A DAG illustrated the hypothesized factors influencing response time and their measurable features. The DAG highlights the variables (in green) that were quantifiable either directly or by combining multiple variables (Fig. 1).

### Feature engineering and utilisation in data analysis

The feature engineering process in this study was designed to extract meaningful insights from a complex dataset comprising over one million EMS missions. By transforming raw data into actionable features, the analysis incorporated diverse factors influencing response

**Table 1** Inclusion and exclusion criteria

Initial dataset N= 1 297 858 observations	
Exclusion criteria	Number of excluded observations
Dispatch priority not 1, 2, or 3 (Transport assignments)	n=480 710
EMS unit type not Emergency Ambulance	
Year not between 2017 and 2022	
Response time not between 0 and 10 h*	
Call handling time not between 0 and 10 h*	
Travel time not between 0 and 10 h*	
On scene time not between 0 and 10 h*	
Transportation time not between 0 and 10 h*	
Delivery time not between 0 and 10 h*	
Intrahospital transports	
<b>Observations included n=817 148 observations</b>	

\* Observations with skewed timestamps caused by technical issues in the prehospital digital platform FRAPP were excluded. Specifically, data with negative time values or response times exceeding 10 h were considered highly unlikely on the basis of the authors' domain expertise and were therefore omitted from the analysis

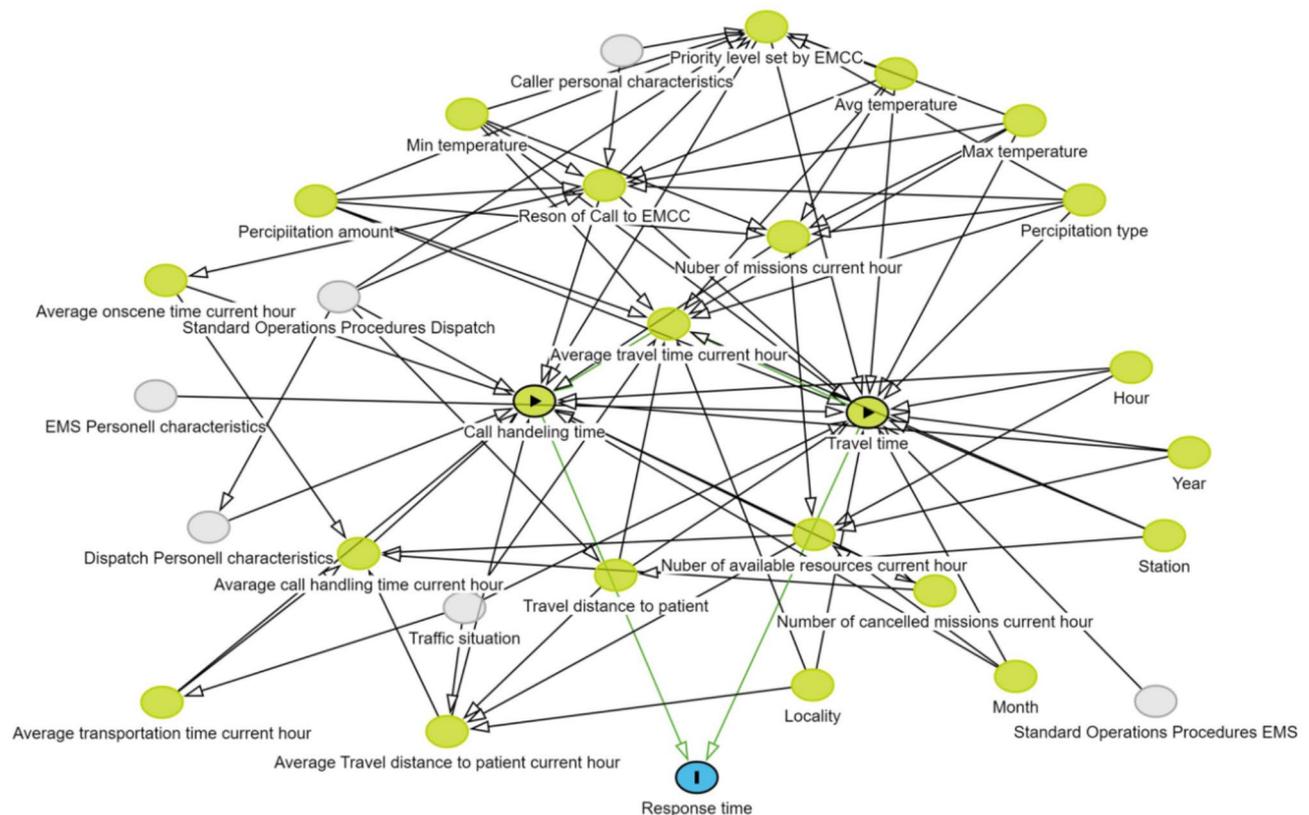
times, including urgency levels, geographical variables, and environmental conditions.

Continuous variables were standardized to a mean of 0 and a standard deviation of 1 to facilitate linear regression [54]. This standardization ensured comparability

across variables with different scales, improving the stability and interpretability of the model. Missing data were incorporated into machine learning models as predictors, acknowledging that missingness itself can contribute valuable insights by reflecting systemic strain or resource availability. For instance, missing timestamps were included as indicators of data flow issues or operational delays.

Notably, the dataset excluded extreme outliers, such as response times exceeding 10 h, which were deemed operationally implausible based on expert knowledge (by object specialist of the data source at Region Stockholm) and technical reviews of timestamp errors. This exclusion ensured that the analysis focused on plausible scenarios while maintaining model reliability.

The features for the machine learning models were designed to align with the specific requirements outlined in the DAG. Key considerations and methodologies included variable grouping, where the original dataset containing 333 categories for call reasons and 115 localities was grouped into 54 and 48 broader categories, respectively. These groupings reduced dimensionality while preserving meaningful distinctions, based on similarities in call type characteristics and geographic



**Fig. 1** Depicts a directed acyclic graph illustrating the relationships between exposures and the outcome, indicating the absence of any open biasing path (60). The features highlighted in green represent variables that were included in the study

proximities to ensure operational relevance and interpretability (Table 2).

Derived features were also created to capture temporal and system-wide dynamics. Hourly averages of response times, call handling times, and travel distances were calculated to reflect system strain during peak periods. Weather-related features, such as temperature and precipitation, were integrated to evaluate their impact on EMS operations. Resource utilization metrics, including the hourly sum of missions and cancellations, provided insights into workload effects.

Additionally, gradient boosting models were used to assess the relative importance of features. Key determinants, such as Emergency Medical Communication Centre (EMCC) priority levels and weather conditions, aligned with prior research and expert judgment. Interaction effects, such as those between locality and adverse weather, were modelled to capture their combined influence on response times, offering a deeper understanding of these interdependencies.

This comprehensive approach ensured that the models effectively represented the complexities of EMS operations, enabling robust predictions and actionable insights.

### Statistical methods

This study employed SAS Enterprise Guide 8.2 for data standardisation and manipulation, ensuring rigorous statistical procedures. Furthermore, SAS Viya 3.05 and Visual Analytics 8.5.2 facilitated comprehensive data analysis. The software allows the user to control the settings and tune the machine learning models. The specific code utilised for these analyses is outlined in Appendix D for easy reference and reproducibility.

### Linear regression

To explore the relationships between the variables and response time, we utilised a linear regression model with 20 bins and a depth of 5 percentiles with a tolerance of  $1e-10$  [53, 54]. The objective of this analysis was to validate the impact of various factors on the response time (Appendix C.2).

### Machine learning models

To construct a reliable prediction model for response time, the dataset underwent partitioning, with 60% allocated for training data, 30% allocated for validation data and 10% as test data. Each model was autotuned by the software to optimise the performance. Various models, including random forest, gradient boosting, neural network, and linear regression, have been investigated and juxtaposed for comparative analysis [52, 53, 61, 62]. Partial Dependence-plots (PDP) were used to interpret the models and gain understanding of the features of interest.

The PDP illustrate the marginal effect of one or two features on the predicted outcome, averaged over the range of values of other features. They help to understand how a single feature influences predictions while holding all other variables constant [63–65]. The gradient boosting model was used to investigate the overall data and data stratified by priority level with the settings reported in Table 3.

### Ethical considerations

All procedures were performed in compliance with relevant laws and institutional guidelines and were approved by the Swedish Ethical Review Authority 2022-09-13 Dnr 2022-03701-01. The study was conducted in accordance with the Declaration of Helsinki and good clinical practice [66, 67].

### Results

The main result showed that complex factors influencing EMS response times. Call handling times impact on overall response time varied by both call reason and urgency level. Travel time and distance likewise differed across call types and priorities. Calls for certain issues (e.g., breathing problems, unconsciousness) consistently had shorter response and travel times, while others (e.g., abdominal pain) had longer handling times, suggesting potential areas for operational improvement (Fig. 2).

### Feature importance

Feature importance varied, with EMCC priority level emerging as the most influential in the overall model (Fig. 3).

Interestingly, some features (e.g., the “hourly sum of missions”) showed opposing effects across different priorities. The result showed an increased workload shortens the Priority 1 response times (perhaps due to resources already are on route to other lower prioritized patients and are redirected to higher prioritized patients), by the PDP used to interpret the model. However, the response time was lengthened for Priority 2 (Fig. 3).

Temperature and precipitation also played roles. Precipitation was generally associated with increased response times, and extreme cold ( $< -8$  °C) affected Priority 1 calls more than moderate temperatures. Conversely, higher temperatures ( $> 18$  °C) tended to affect lower-priority calls. Local geography influenced response time variations, particularly in how urgency levels interacted with different areas. PDPs revealed that high-priority calls in central Stockholm had consistently lower response times compared to suburban or rural areas. However, for lower-priority calls, this trend reversed in some cases, with central areas experiencing longer response times than expected. This suggests that resource saturation at peak times may lead to a trade-off

**Table 2** Explanation of crafted feature

Feature	Explanation
Priority level set by EMCC	The priority level set by EMCC is presumed to affect response time due to the assessed urgency. Since all missions in all priority levels affect response time between priority levels, stratification between priority levels is not suitable. Instead, the priority level is handled as a feature affecting response time like the other variables investigated in this study.
Response time	A created feature using the time difference between when the EMCC call was received, and the first EMS resource arrival at scene of incident. The response time is a result of call handling time and travel time.
Call Handling time	A created feature using the time difference between when the EMCC call was received, and the first EMS resource received a mission from EMCC in the incident.
Travel time	A created feature using the time difference between when the first EMS resource received a mission from EMCC in the incident, and the first EMS resource arrival at scene of incident.
Reason of Call to EMCC	The reason of call to EMCC was grouped from the origin of 333 different categories, which differs slightly between different years, into 54 different categories reflecting the reason of call to EMCC. Reason of Call to EMCC is presumed to effect call handling time because differences in call handling time due to the nature of different reason of call to EMCC-categories. Reason of Call to EMCC is presumed to effect travel time because differences in travel time in different reason of call to EMCC-categories where personal EMS-personnel experience might influence travel time.
Locality	The locality was grouped from the origin of 115 localities, where the same localities was duplicated one or several times due to misspelling, into 48 localities which reflect the real locality. Locality is presumed to affect call handling time due to amount of available resources might differ in different localities. If there is lack of resources in a locality the call handling time might increase due to finding nearest suitable resource. Locality is presumed to affect travel distance and travel time because different coverage of available resources in different localities. Differences related to locality might indicate EMS allocation out of tune.
Hour	The hour of day is used to investigate differences in response time between hour of day. Differences related to hour might indicate EMS allocation out of tune.
Month	The month of year is used to investigate differences between month of year. Differences related to month might indicate EMS allocation out of tune.
Year	The year is used to identify differences in years like the Covid-19 pandemic and differences in EMS allocation between years. Differences between years might indicate EMS allocation out of tune.
On scene time	A created feature using the time difference between when the first EMS resource arrival at scene of incident and the first EMS resource started transportation of patient.
Transportation time	A created feature using the time difference between when the first EMS resource started transportation of patient and the first EMS resource arrival at hospital.
Air temperature	The average temperature current day at the weather station called "Observatoriekullen" in the centre of Stockholm. Differences in temperature is presumed to affect response time in different ways, like low temperature in combination with precipitation could affect traffic and travel time. It could also generate differences in reason of call to EMCC related to high and low temperatures.
Airtemperature_max	The max temperature current day at the weather station called "Observatoriekullen" in the centre of Stockholm. Differences in max temperature is presumed to affect response time in different ways, like low temperature in combination with precipitation could affect traffic and travel time. It could also generate differences in reason of call to EMCC related to high and low temperatures.
Airtemperature_min	The min temperature current day at the weather station called "Observatoriekullen" in the centre of Stockholm. Differences in min temperature is presumed to affect response time in different ways, like low temperature in combination with precipitation could affect traffic and travel time. It could also generate differences in reason of call to EMCC related to high and low temperatures.
H_AVG_of_Responsetime	A created feature aggregating the response time as average response time of all missions current hour. The purpose is to reflect how the general response time the current hour effects response time in individual missions. Differences in the hourly average response time is presumed to reflect the system where increased hourly response time might indicate strained situation or EMS allocation out of tune.
H_AVG_of_Call_Handlingtime	A created feature aggregating the call handling time as average response time of all missions current hour. The purpose is to reflect how the general call handling time the current hour effects response time in individual missions. Differences in the hourly average call handling time is presumed to reflect the system where increased hourly call handling time might indicate strained situation or EMS allocation out of tune.
H_AVG_of_DistArrival_km	A created feature aggregating the travel distance to patient in all missions current hour. The purpose is to reflect how the general travel distance to patient the current hour effects response time in individual missions. Differences in the hourly average travel distance is presumed to reflect the system where increased travel distance might indicate strained situation or EMS allocation out of tune.

**Table 2** (continued)

Feature	Explanation
H_AVG_of_Drivetime	A created feature aggregating the drive time as average drive time of all missions current hour. The purpose is to reflect how the general drive time the current hour effects response time in individual missions. Differences in the drive time is presumed to reflect the system where increased travel distance might indicate strained situation or EMS allocation out of tune. The traffic situation is a factor highly affecting drive time.
H_AVG_of_Onscenetime	A created feature aggregating the on scene time as average on scene time of all missions current hour. The purpose is to reflect how the general on scene time the current hour effects response time in individual missions. Differences in the on scene time is presumed to indicate prolonged treatment time affecting the response time.
H_AVG_of_Transportationtime	A created feature aggregating the on transportation time as average transportation time of all missions current hour. The purpose is to reflect how the general transportation time the current hour effects response time in individual missions. Differences in the on scene time is presumed to indicate prolonged transportation time affecting the response time.
H_SumOfTasks	A created feature aggregating the sum of missions current hour. The purpose is to reflect how the sum of missions the current hour effects response time in individual missions. Differences in the sum of missions is presumed in relation to available resources is presumed affecting the response time.
H_Sum_of_MissionCancelled	A created feature aggregating the sum of cancelled missions the current hour. The purpose is to reflect how the sum of cancelled missions the current hour effects response time in individual missions. Differences in the sum of cancelled missions is presumed to reflect the system, where a strained situation with lack of resources, or EMS allocation out of tune, would give a higher rate of cancelled missions and affect the response time.
H_C_DIST_of_rakelid	A created feature aggregating the sum of distinct EMS resource ID the current hour. The purpose of this variable is to reflect how the sum of EMS resources assigned to a mission the current hour effects response time in individual missions. Differences in the sum of EMS resources in relation to the sum of missions affects response time.

**Table 3** Gradient boosting settings

Gradient Boosting model settings				
	Overall	Priority 1	Priority 2	Priority 3
Auto-stop method	Stagnation	Stagnation	Stagnation	Stagnation
Auto-stop iterations	4	4	4	4
Number of trees	50	50	50	50
Learning rate	0.23	0.23	0.23	0.23
Subsample rate	0.1	1	1	1
Lasso	6.11111111	1.11111111	1.11111111	0
Ridge	2.22222222	6	1	2.22222222

where high-priority cases are prioritized at the expense of lower-priority calls. Additionally, in some lower-density areas, ambulances may be stationed farther apart, leading to naturally longer response times, which disproportionately affect lower-priority dispatches (Appendix A.34).

The results show that different features have different effect on response time which confirms the relationships in the DAG (Fig. 1). Multiple features interact to influence EMS response time. PDPs show that while urgency level is the strongest predictor of response time, its effect varies depending on geographic location and real-time system strain. For instance, in areas with high EMS demand, the travel time component increases significantly for lower-priority calls, whereas high-priority calls maintain stable response times. This effect suggests that EMS resource allocation strategies should not only

prioritize high-urgency cases but also account for cumulative system strain to prevent excessive delays in less urgent cases.

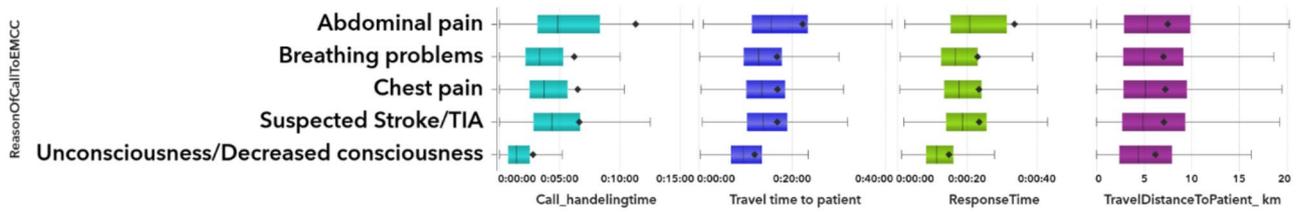
One example is determinants driving calls to EMCC, including call handling duration, travel time to the patient, and the urgency level of the situation, which collectively influence response time (Fig. 2). Among these factors, both the duration of call handling and the travel time to the patient emerge as influencers alongside the priority level set by EMCC, introducing variability into the overall response time.

A PDP revealed that specific reasons for contacting the EMCC, such as breathing problems and loss of consciousness, are associated with shorter call handling and travel times. This finding underscores that these variables not only directly impact response time but also exert indirect effects by influencing other factors. For example, within priority level 1, variations in response time are observed across different reasons for contacting the EMCC (Appendix A.18). The predicted response time differs from about 15 min for Unconsciousness and Cardiac arrest compared to over 30 min for Abdominal pain and Extremity/Wound/Minor trauma, all in the same priority level 1 (Fig. 4).

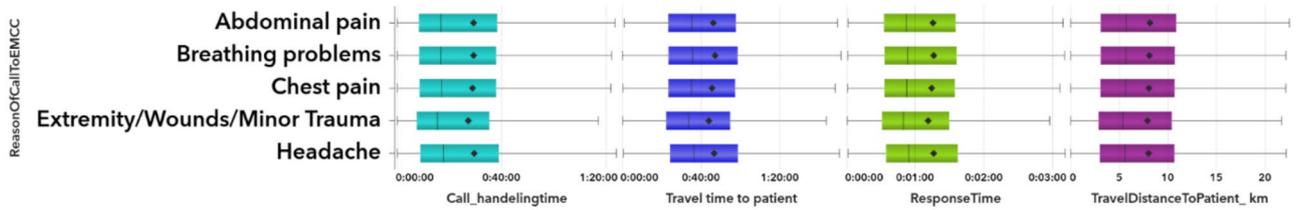
By shedding light on these relationships, this study elucidates how various factors interact to shape emergency response times, offering insights crucial for optimising EMS on the basis of response time.

The EMS response time is subject to the influence of various factors beyond immediate call handling and travel logistics. Among these factors, the timing of incidents throughout the day is a significant determinant.

Priority 1



Priority 2



Priority 3

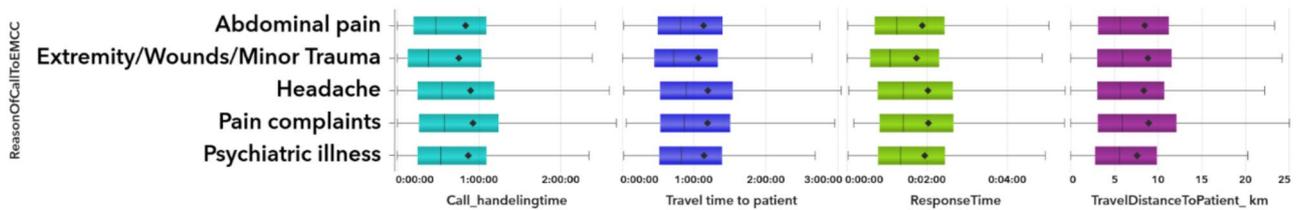
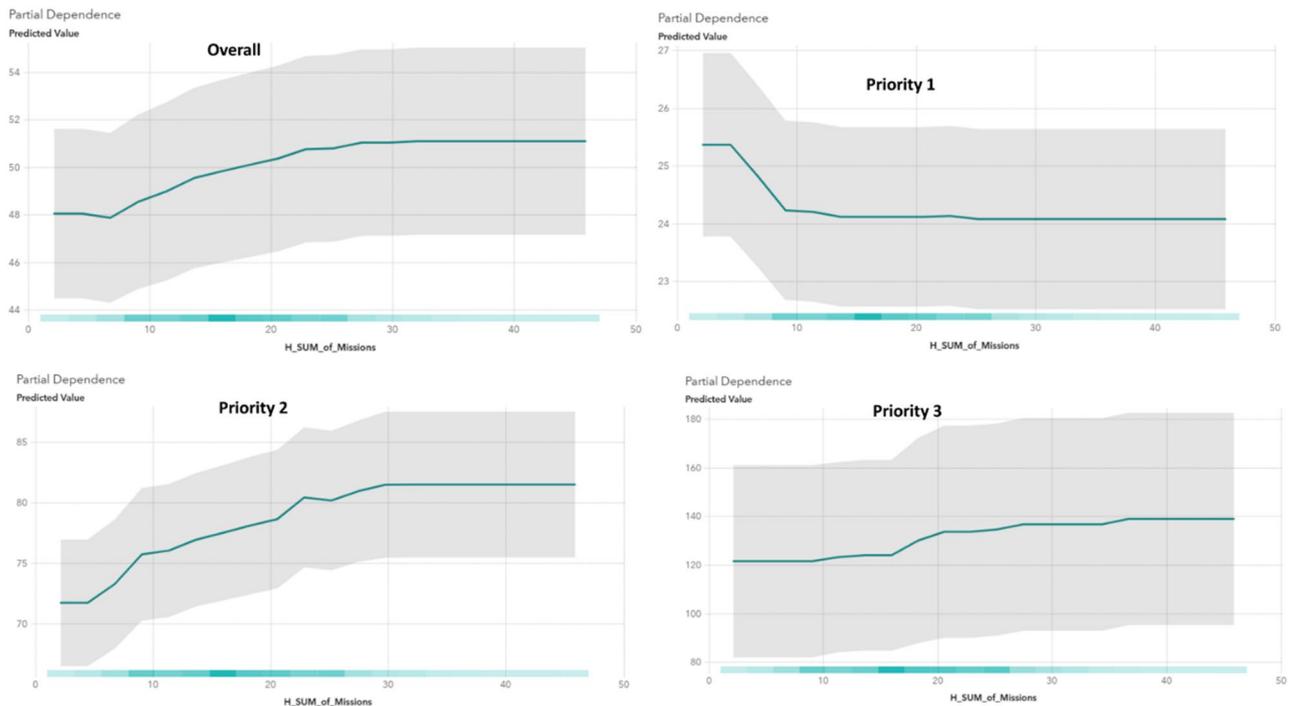


Fig. 2 Boxplots of call handling, travel, and response times for the 5 most common call types by EMCC priority

Variable	Overall Importance	Priority 1 Importance	Priority 2 Importance	Priority 3 Importance
Priority set by EMCC	20731343.33	0.00	0.00	0.00
H_AVG_of_Responsetime	7185332.20	145456.79	9526485.43	2395520.23
H_AVG_of_Call_handlingtime	6556164.58	137597.84	3067662.84	744410.27
TravelDistanceToPatient_km	2515488.65	1003754.09	1836551.47	130192.23
EMS Station	1410885.77	151454.27	1344546.81	347773.08
Hour	493230.60	27145.80	595903.08	75136.52
H_AVG_of_Drivetime	437582.35	279663.73	168720.13	100102.51
ReasonOfCallToEMCC	421833.39	374796.38	316705.51	42494.85
Year	241493.44	35423.51	261433.72	56641.28
Locality	211631.92	34908.77	237804.73	48728.20
H_SUM_of_Missions	93263.67	9180.96	100008.70	51450.61
H_AVG_of_DistArrvalkm	69461.08	27658.79	79973.62	26651.01
Weekday	69229.23	24204.39	120351.64	50996.07
Airtemperature_max	58071.24	37844.49	49879.85	12265.03
H_AVG_of_Deliverytime	42691.20	9765.33	55577.85	35536.29
Month	42187.22	5346.95	58484.90	8957.62
Airtemperature_min	41984.89	11127.81	33140.05	10967.13
H_AVG_of_Transportationtime	41603.74	19424.88	57738.28	25929.32
Airtemperature	39089.81	12618.02	39433.94	17837.34
H_AVG_of_Onscenetime	34331.05	22383.99	61492.07	19483.43
H_SUM_of_Resources	31664.93	6331.16	21128.70	23911.55
AmmountPrecipitation	31581.42	16847.88	52545.71	16913.67
PrecipitationType	26412.15	11572.71	45536.67	16086.03
H_SUM_of_MissionCanc	0.00	601.59	1325.87	1547.23

Fig. 3 Feature importance in overall data and by urgency level; color-coding to show highest impact



**Fig. 4** PDP of the hourly total sum of missions at different priority levels compared to predicted response time

The analysis revealed discernible fluctuations in response times across different hours, with particularly noteworthy impacts evident during periods of varying urgency levels, especially compared with instances of the highest urgency (Appendix A.22).

Furthermore, the geographical location of incidents plays a pivotal role in shaping response time dynamics across different urgency levels. Notably, the central region of Stockholm has markedly shorter response times during high-urgency scenarios. Conversely, during lower-urgency scenarios, this same area experiences prolonged response times in comparison to other geographical localities (Appendix A.34).

#### Distribution of response time

Response time distribution was heavily skewed for priority levels 2 and 3, featuring a “long tail” of extended response times (Fig. 5). This skew can reduce ML model generalizability by overemphasizing outliers. Because the aim was to deepen the understanding of the complex factors influencing EMS response times, including these long-tail cases, rather than to build a perfect predictive model, this trade-off was considered acceptable (see Fig. 6).

#### Model performance

Overall, the gradient boosting model produced an ASE of 1757.70 min in the test data. Performance improved when stratified by priority level, with Priority 1 yielding the best prediction accuracy (ASE = 545.99) and Priority

3 the lowest (ASE = 6752.75), reflecting that Priority 3 data were more skewed (Table 4).

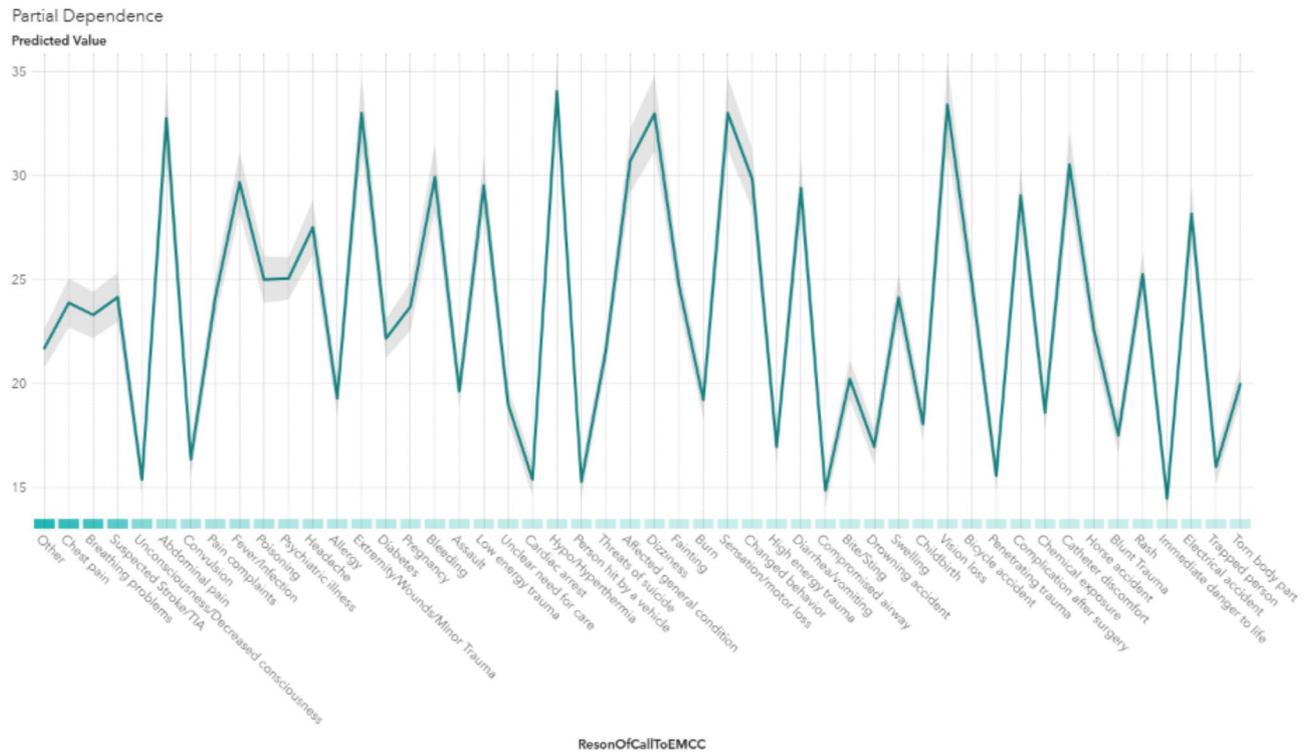
Linear regression indicated significant associations between a vast number of variables and response time (average squared error [ASE] = 2127.1291 in the test data;  $R^2 = 0.3706$ ). This  $R^2$  value, while not extremely high, aligns with the DAG’s suggestion that additional unmeasured factors also play a role. Of the tested ML models, gradient boosting demonstrated the best fit, though differences among models were slight. Table 5 (below) compares model performance (Appendix B.2).

## Discussion

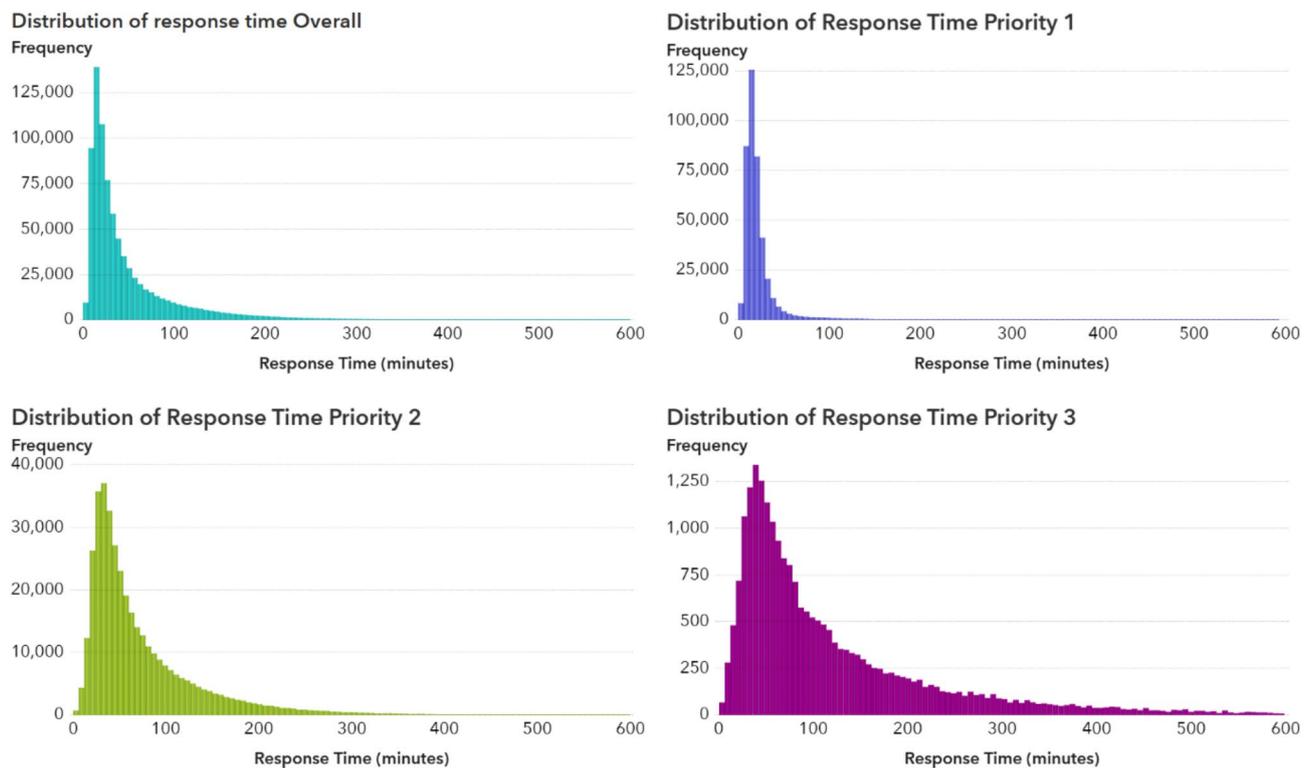
### Key results

This study aimed to deepen the understanding of the complex factors influencing EMS response times. Confirming previous research [9, 10], call handling and travel time are vital, but new detail emerged on how different call types within the same priority can yield substantial variations. Calls for unconsciousness produced shorter handling times, while abdominal pain or urinary issues were lengthier all in the highest priority level.

Environmental factors like precipitation also significantly influenced response times, aligning with prior findings [33, 35, 40, 42]. The Machine Learning model allowed for better recognition of these “uncontrollable” contributors to response time.



**Fig. 5** PDP of the Reason of Call to EMCC compared to predicted response time in the priority 1 level. Calls for abdominal pain and minor trauma experience longer response times despite being priority 1 while unconsciousness and cardiac arrest exhibit the shortest response times



**Fig. 6** The histogram shows distribution of response time in different priority levels set by EMCC

**Table 4** Gradient boosting machine learning models, overall and stratified by priority level set by EMCC

Partition	ASE	Observed Average	SSE	Observations Used	Unused
Overall Training	1677.74	173.19	822579900.75	490289	0
Overall Validation	1778.43	169.56	435970851.38	245144	0
Overall Test	1757.70	170.25	143630059.64	81715	0
Priority 1 Training	528.72	58.65	130357237.02	246553	0
Priority 1 Validation	563.73	51.59	69360839.96	123040	0
Priority 1 Test	545.99	53.12	22416336.13	41056	0
Priority 2 Training	2544.84	192.32	585570273.01	230101	0
Priority 2 Validation	2771.18	179.75	319591914.52	115327	0
Priority 2 Test	2750.57	182.03	105528536.65	38366	0
Priority 3 Training	5442.37	350.41	74206757.86	13635	0
Priority 3 Validation	6832.91	309.86	46306608.17	6777	0
Priority 3 Test	6752.75	306.42	15484065.72	2293	0

**Table 5** Model comparison where gradient boosting was selected as the best fit

Selected	Number Of Observations	Percentile	Predicted Average	ASE	Observed Average	SSE	Model
Yes	490,289	5	167.80	1690.09	172.90	828634299.32	Training: Gradient boosting
Yes	245,144	5	168.22	1789.91	169.40	438785173.01	Validation: Gradient boosting
Yes	81,715	5	167.29	1769.54	170.06	144598368.52	Test: Gradient boosting
No	490,289	5	168.66	1307.06	194.87	640835098.42	Training: Forest
No	245,144	5	166.32	1796.18	168.69	440322314.81	Validation: Forest
No	81,715	5	165.52	1786.66	169.21	145997039.91	Test: Forest
No	222,482	5	167.41	1788.88	169.01	397994152.56	Training: Neural network
No	110,779	5	168.11	1861.60	166.61	206226675.73	Validation: Neural network
No	37,213	5	168.51	1832.01	166.76	68174404.09	Test: Neural network
No	222,482	5	129.17	2116.45	158.05	470871881.72	Training: Linear regression
No	110,787	5	129.46	2138.34	157.61	236900262.60	Validation: Linear regression
No	37,214	5	129.65	2127.13	157.27	79158981.29	Test: Linear regression

**Implications for EMS management**

By leveraging machine learning for predictive modeling, EMS agencies can optimize resource allocation and improve response times. To operationalize these insights, EMS should develop a dynamic resource allocation system by implementing real-time monitoring dashboards that integrate machine-learning-driven predictive alerts for demand surges based on call type, weather conditions, and geographic trends. Shifting from fixed stationing to a fluid deployment model, where ambulances dynamically reposition based on real-time forecasts, would enhance efficiency.

Emergency call prioritization can be improved by deploying AI-assisted triage systems in Emergency Medical Communication Centers (EMCCs) to refine priority classification and reduce misallocations of high-priority responses. Training dispatchers to use machine learning-backed decision support tools would ensure that emergency calls are matched with appropriate response levels based on historical trends and real-time data. Integrating

traffic and weather data for smarter routing is also essential. Automated rerouting protocols should be established to adjust dispatch recommendations based on live traffic feeds and adverse weather forecasts, while partnerships with municipal traffic authorities could facilitate the creation of EMS priority lanes or dynamic traffic signal preemption for ambulances in high-demand areas.

Addressing equity in response times requires a geospatial analysis to identify high-risk zones with prolonged response times and adjust resource distribution accordingly. Allocating additional EMS units or mobile response teams in underserved areas during peak demand periods or severe weather conditions would ensure a more balanced deployment of resources. Furthermore, improving data-driven decision-making through training and feedback mechanisms is crucial. Specialized training programs should be developed to ensure EMS personnel and dispatchers can interpret machine learning-based insights and adjust operational decisions accordingly. Establishing a feedback mechanism where EMS field

teams report real-world response discrepancies would enable continuous model refinement for improved accuracy.

Strengthening collaboration between EMS and policymakers is necessary to translate these insights into structural improvements. Engaging local governments in policy discussions on urban planning would help ensure that new infrastructure investments align with machine learning-based response efficiency insights. Additionally, advocating for funding support to scale real-time machine learning deployment, particularly in regions with historically slow response times, could facilitate the implementation of these advanced strategies.

By systematically adopting these measures, EMS organizations can transition from reactive response models to proactive, data-driven emergency care strategies. These changes will enhance response efficiency, improve equity in service delivery, and ultimately lead to better patient outcomes.

#### **Methodological considerations**

This study employed a robust methodological framework combining advanced machine learning techniques and traditional statistical approaches to analyze a complex dataset of over one million EMS missions. A key strength of the methodology was the use of feature engineering and the construction of a directed acyclic graph (DAG) to elucidate relationships among variables. This approach ensured a comprehensive understanding of the factors influencing response times.

The dataset's partitioning into training, validation, and test sets allowed for rigorous model evaluation, and the use of Gradient Boosting models provided nuanced insights into feature importance. However, the exclusion of extreme outliers, such as response times exceeding 10 h, may limit the generalizability of the findings to atypical cases.

Ethical considerations were integral to the study's design, with adherence to the STROBE checklist ensuring clarity and rigor. Limitations include the reliance on aggregate metrics, which may obscure finer temporal trends. The absence of traffic data and other operational constraints also highlights areas for future research.

#### **Challenges and considerations for real-time implementation**

The real-time implementation ML predictions in EMS operations presents both significant opportunities and notable challenges. While the potential of ML to dynamically optimize resource allocation is clear, translating these insights into live operations requires addressing several practical considerations.

A key requirement for real-time deployment is robust computational infrastructure. Systems capable of

processing large datasets and generating instantaneous predictions are essential. This may necessitate investments in advanced technologies such as cloud-based solutions or edge computing to meet the demanding computational requirements.

Seamless integration of diverse data sources is another critical factor for effective real-time predictions. Combining weather updates, traffic conditions, and real-time EMS resource availability requires interoperable systems and standardized data protocols to ensure smooth operation and data flow.

Operational constraints within EMS systems must also be carefully considered. Regulatory and logistical requirements, including variations in response protocols, resource limitations, and the need for rapid decision-making during high-stakes scenarios, necessitate models that are both adaptable and compliant with these constraints.

The adoption of real-time ML systems further depends on adequate training for EMS personnel and dispatchers. These individuals must be equipped to interpret ML predictions and incorporate them into their workflows. User-friendly interfaces and decision support tools can significantly enhance the usability and acceptance of these systems.

Lastly, continuous monitoring and feedback mechanisms are essential for maintaining the accuracy and reliability of real-time ML systems. Establishing feedback loops that refine models based on operational outcomes will ensure their long-term utility and trustworthiness.

By addressing these challenges, the implementation of real-time ML systems can transition from theoretical potential to practical reality, enhancing the efficiency and responsiveness of EMS operations.

#### **Limitations and directions for future research**

The study identifies several limitations, including the exclusion of traffic data and operational constraints, but further elaboration is necessary to contextualize their impact and provide direction for future research. Traffic conditions are a critical factor influencing EMS response times, especially in urban environments. The absence of traffic data in the models could result in underestimating delays during periods of peak congestion. Future studies should aim to incorporate real-time traffic feeds or historical congestion patterns to enhance model accuracy.

Operational variability also poses challenges to the generalizability of the findings. Differences in EMS standard operating procedures, such as dispatch protocols and resource allocation strategies, could influence outcomes. Expanding the analysis to include regional or organizational variations would provide a more comprehensive understanding of these effects and their implications.

While the study includes weather data, its focus is limited to Stockholm. Regional weather patterns, such as extreme heat in arid climates or heavy snowfall in colder regions, may have distinct impacts on response times. Exploring these variations could improve the robustness and applicability of the models to diverse geographic settings.

The use of aggregated features, such as hourly averages, may obscure finer temporal trends. To address this, future research should investigate shorter time intervals to better capture fluctuations in EMS demand and response dynamics. This approach would allow for a more detailed understanding of system behaviour under varying conditions.

Finally, while the study highlights the potential for real-time machine learning applications, it does not address the technical and logistical challenges of implementing such systems in live EMS operations. Investigating successful case studies of real-time ML integration could provide valuable insights and practical solutions for overcoming these barriers.

By addressing these limitations, future research can build on the current findings to further advance the effectiveness and reliability of EMS resource planning and response strategies.

### Generalizability

The machine learning model developed in this study demonstrates a high level of adaptability for other EMS systems, as it is built on widely applicable factors such as call handling time, travel distance, urgency levels, weather conditions, and geographic variability. These determinants influence EMS response times universally, making the model broadly relevant to emergency medical services beyond Stockholm. The use of Gradient Boosting and feature engineering techniques ensures that the model captures complex interactions between these variables, allowing for meaningful insights into response time dynamics. Given that many urban EMS systems operate with structured dispatch protocols and centralized emergency communication centres, the core structure of the model can be readily applied in similar settings.

However, certain modifications would be necessary when adapting the model to EMS systems with different operational structures. Some EMS systems categorize priority levels differently or use alternative dispatch methods, which may require adjustments to the model's classification framework. Additionally, EMS systems that utilize a tiered response structure, where basic life support and advanced life support teams are dispatched separately, may need to introduce additional features to capture these nuances in resource allocation.

Geographic and traffic-related variations also present challenges when applying the model across different

regions. While the Stockholm EMS system operates in a metropolitan setting with both fixed and fluid deployment strategies, the model may require recalibration when applied to rural areas where response times are naturally longer due to increased distances between incidents and available EMS units. Furthermore, the absence of real-time traffic data in the Stockholm dataset means that the model may need to be supplemented with live traffic inputs in cities where congestion significantly impacts response times.

### Conclusion

This study provides a comprehensive analysis of the multifactorial determinants of EMS response times, demonstrating the significant influence of variables such as urgency levels, weather conditions, and resource availability. By leveraging advanced machine learning techniques, particularly Gradient Boosting models, the research offers a nuanced understanding of how these factors interact and provides a robust framework for predictive analytics in EMS operations.

Key findings underscore the importance of incorporating real-time data into dynamic resource allocation strategies. For high-priority calls, optimizing dispatch processes and leveraging predictive models can reduce response times and enhance patient outcomes. For lower-priority calls, addressing systemic inefficiencies, such as resource imbalances and weather-related delays, can improve equity and service delivery.

The implications for EMS management are substantial. Transitioning from static to adaptive deployment models can enable agencies to better respond to fluctuations in demand and external conditions. Integrating machine learning predictions into live operations offers a path to more efficient, equitable, and sustainable emergency care systems.

Future research should address limitations such as the exclusion of traffic data and operational constraints and explore the generalizability of these models to rural or smaller urban settings. Policymakers and EMS administrators are encouraged to consider the study's findings to inform decisions on funding, resource allocation, and deployment strategies.

By bridging the gap between academic research and operational practice, this study lays the groundwork for a data-driven evolution in EMS planning, aiming to optimize response times and improve outcomes for all patients, regardless of urgency level or location.

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12911-025-02975-z>.

Supplementary Material 1

## Acknowledgements

We like to express our gratitude to Mathias Lanner at SAS Institute for his support and guidance regarding the use of SAS software in this study.

## Author contributions

P.H. Conceptualization, Methodology, Formal analysis, Investigation, Data Curation, Writing - Original Draft, Visualization. J.L. Conceptualization, Methodology, Supervision. D.J. Conceptualization, Methodology, Supervision. P.B. Conceptualization, Methodology. V.V. Conceptualization, Methodology, Supervision, Project administration.

## Funding

Open access funding provided by Karolinska Institute.

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

## Data availability

Due to the sensitive nature of the used data in this study, where data is confidential due to the Swedish Public Access to Information and Secrecy Act. It is possible to request the used data from Region Stockholm and their Centre of Health data.

## Declarations

### Ethics approval and consent to participate

This study was conducted in full compliance with applicable laws and institutional guidelines, with ethical approval granted by the Swedish Ethical Review Authority on 13 September 2022 (Dnr 2022-03701-01). The Swedish Ethical Review Authority explicitly exempted and waived the requirement for individual consent, in accordance with the Swedish Law (SFS 2003:460) on the Ethical Review of Research Involving Humans, which allows for research without consent under specified conditions like register studies.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

Received: 10 October 2024 / Accepted: 17 March 2025

Published online: 24 March 2025

## References

1. Ambulance care in Europe - Organization and practices of ambulance services in 14 European countries. NIVEL; 2015.
2. Alshehri MF, Pigoga JL, Wallis LA. A mixed methods investigation of emergency communications centre triage in the government emergency medical services system, cape town, South Africa. *Afr J Emerg Medicine: Revue Africaine De La Médecine D'urgence*. 2020;10(Suppl 1):S12–7. <https://doi.org/10.1016/j.afjem.2020.02.004>.
3. SOSFS. 2009:10 Socialstyrelsens föreskrifter om ambulanssjukvård m.m. [SOSFS 2009:10 The National Board of Health and Welfare's regulations on ambulance medical care, etc.]. In: Socialstyrelsen, ed.: Socialstyrelsen, 2020.
4. Bohm K, Kurland L. The accuracy of medical dispatch - a systematic review. *Scand J Trauma Resusc Emerg Med*. 2018;26(1):94. <https://doi.org/10.1186/s13049-018-0528-8>.
5. Castrén M, Kurland L, Liljegård S, et al. Non-specific complaints in the ambulance; predisposing structural factors. *BMC Emerg Med*. 2015;15:8. <https://doi.org/10.1186/s12873-015-0034-5>.
6. Ceklic E, Tohira H, Ball S, et al. A predictive ambulance dispatch algorithm to the scene of a motor vehicle crash: the search for optimal over and under triage rates. *BMC Emerg Med*. 2022;22(1):74. <https://doi.org/10.1186/s12873-022-00609-5>.
7. Holmström IK, Kaminsky E, Lindberg Y, et al. The perspectives of Swedish registered nurses about managing difficult calls to emergency medical dispatch centres: a qualitative descriptive study. *BMC Nurs*. 2021;20(1):150. <https://doi.org/10.1186/s12912-021-00657-5>.
8. Salminen T, Kaartinen K, Roos M, et al. Differences between the dispatch priority assessments of emergency medical dispatchers and emergency medical services: a prospective register-based study in Finland. *Scand J Trauma Resusc Emerg Med*. 2023;31(1):8. <https://doi.org/10.1186/s13049-023-01072-2>.
9. Majlesinasab N, Maleki M, Nikbakhs E. Performance evaluation of an EMS system using queuing theory and location analysis: A case study. *Am J Emerg Med*. 2022;51:32–45. <https://doi.org/10.1016/j.ajem.2021.10.004>.
10. Cabral ELDS, Castro WRS, Florentino DRM, et al. Response time in the emergency services. Systematic review. *Acta Cirúrgica Brasileira*. / 2018;33(12):1110–21. <https://doi.org/10.1590/s0102-865020180120000009>.
11. Heidet M, Da Cunha T, Brami E, et al. EMS access constraints and response time delays for deprived critically ill patients near Paris, France. *Health Aff (Millwood)*. 2020;39(7):1175–84. <https://doi.org/10.1377/hlthaff.2019.00842>.
12. Park JH, Song KJ, Shin SD, et al. Does second EMS unit response time affect outcomes of OHCA in multi-tiered system? A nationwide observational study. *Am J Emerg Med*. 2021;42:161–67. <https://doi.org/10.1016/j.ajem.2020.02.018>.
13. Responstid för ambulans vid prio 1-larm <https://vardenisiffor.se>: Swedish Association of Local Authorities and Regions. 2023 [updated 2021-22-26; cited 2024-04-15]. Available from: <https://vardenisiffor.se/indikator/0ff6e916-ad43-43e0-bbfc-5ce18d510a9e?datefrom=2016-01-01%26dateto=2020-12-31%26periodtype=year%26relatedmeasuresbyentry=kallsystem&relatedmeasuresbyid=f65721b7-3be6-423e-9ddf-b52381caa76f&units=01> accessed 2024-04-15 2024.
14. Thomas S, Foley C, Kane B, et al. Variation in resource allocation in urgent and emergency care systems in Ireland. *BMC Health Serv Res*. 2019;19(1):NPAG–NPAG. <https://doi.org/10.1186/s12913-019-4504-4>.
15. Umam MIH, Santosa B, Siswanto N. A simulation optimization for location and allocation of emergency medical service. *Int J Online Biomedical Eng*. 2022;18(11):158–72. <https://doi.org/10.3991/ijoe.v18i11.31055>.
16. Wen TX, Zhang ZN, Qiu M, et al. A Multi-Objective optimization method for emergency medical resources allocation. *J Med Imaging Health Inf*. 2017;7(2):393–99. <https://doi.org/10.1166/jmih.2017.2027>.
17. Yang W, Su Q, Zhou M, et al. Ambulance allocation considering the Spatial randomness of demand. *Comput Ind Eng*. 2020;139. <https://doi.org/10.1016/j.cie.2019.106202>.
18. Yoon S, Albert LA. An expected coverage model with a cutoff priority queue. *Health Care Manag Sci*. 2018;21(4):517–33. <https://doi.org/10.1007/s10729-017-9409-3>. [published Online First: 2017/07/20].
19. Yuan YF, Song QK, Zhou B. A Wasserstein distributionally robust chance constrained programming approach for emergency medical system planning problem. *Int J Syst Sci*. 2022;53(10):2136–48. <https://doi.org/10.1080/00207721.2022.2040641>.
20. Yuangyai C, Nilsang S, Cheng CY. Robust ambulance base allocation strategy with social media and traffic congestion information. *J Ambient Intell Humaniz Comput* 2020 <https://doi.org/10.1007/s12652-020-01889-0>
21. Yunus S, Abdulkarim IA. Road traffic crashes and emergency response optimization: a geo-spatial analysis using closest facility and location-allocation methods. *Geomatics Nat Hazards Risk*. 2022;13(1):1535–55. <https://doi.org/10.1080/19475705.2022.2086829>.
22. Zaheeruddin GH. Optimally allocating ambulances in Delhi using mutation based shuffled frog leaping algorithm. *Int J Adv Comput Sci Appl*. 2022;13(9):364–74. <https://doi.org/10.14569/ijacsa.2022.0130942>.
23. Zhang R, Zeng B. Ambulance deployment with relocation through robust optimization. *IEEE Trans Autom Sci Eng*. 2019;16(1):138–47. <https://doi.org/10.1109/tase.2018.2859349>.
24. Zhang ZH, Li K. A novel probabilistic formulation for locating and sizing emergency medical service stations. *Ann Oper Res*. 2015;229(1):813–35. <https://doi.org/10.1007/s10479-014-1758-4>.
25. Zhou L, Wang SH, Xu ZB. A Multi-factor Spatial optimization approach for emergency medical facilities in Beijing. *Isprs Int J Geo-Information*. 2020;9(6). <https://doi.org/10.3390/ijgi9060361>.
26. Zhou ZY, Matteson DS, PREDICTING MELBOURNE AMBULANCE DEMAND USING, KERNEL WARPING. *Annals Appl Stat*. 2016;10(4):1977–96. <https://doi.org/10.1214/16-aos961>.
27. Zhou ZY, Matteson DS, Woodard DB, et al. A Spatio-Temporal point process model for ambulance demand. *J Am Stat Assoc*. 2015;110(509):6–15. <https://doi.org/10.1080/01621459.2014.941466>.
28. Zonouzi MN, Kargari M. Modeling uncertainties based on data mining approach in emergency service resource allocation. *Comput Ind Eng*. 2020;145. <https://doi.org/10.1016/j.cie.2020.106485>.

29. Zou Y, Jia L, Chen SJ, et al. Spatial accessibility of emergency medical services in Chongqing, Southwest China. *Front Public Health*. 2023;10. <https://doi.org/10.3389/fpubh.2022.959314>.
30. Murray B, Kue R. The use of emergency lights and sirens by ambulances and their effect on patient outcomes and public safety: A comprehensive review of the literature. *Prehosp Disaster Med*. 2017;32(2):209–16. <https://doi.org/10.1017/S1049023X16001503>.
31. Torres N, Trujillo L, Maldonado Y, et al. Correction of the travel time Estimation for ambulances of the red cross Tijuana using machine learning. *Comput Biol Med*. 2021;137:104798. <https://doi.org/10.1016/j.combiomed.2021.104798>. [published Online First: 2021/09/06].
32. Viglino D, Vesin A, Ruckly S, et al. Daily volume of cases in emergency call centers: construction and validation of a predictive model. *Scand J Trauma Resusc Emerg Med*. 2017;25(1):86. <https://doi.org/10.1186/s13049-017-0430-9>. [published Online First: 2017/08/31].
33. Chen F, Yip PSF. The influence of cold weather on the usage of emergency link calls: a case study in Hong Kong. *BMC Med Inf Decis Mak*. 2015;15. <https://doi.org/10.1186/s12911-015-0191-1>.
34. Hanefeld C, Klaassen-Mielke R, Miebach J, et al. Influence of extreme weather conditions on the deployment volume of emergency medical services. *Medizinische Klinik-Intensivmedizin Und Notfallmedizin*. 2021;116(2):154–60. <https://doi.org/10.1007/s00063-019-00641-7>.
35. Karamercan MA, Dündar ZD, Ergin M, et al. Seasonal variations of patients presenting dyspnea to emergency departments in Europe: results from the EURODEM study. *Turk J Med Sci*. 2020;50(8):1879–86. <https://doi.org/10.3906/sag-2002-221>.
36. Nishimura T, Rashed EA, Kodera S, et al. Social implementation and intervention with estimated morbidity of heat-related illnesses from weather data: A case study from Nagoya City. *Japan Sustainable Cities Soc*. 2021;74. <https://doi.org/10.1016/j.scs.2021.103203>.
37. Oka K, Hijioka Y. Prediction of the number of heatstroke patients transported by ambulance in Japan's 47 prefectures: proposal of heat acclimatization consideration. *Environ Res Commun*. 2021;3(12). <https://doi.org/10.1088/2515-7620/ac3d21>.
38. Ramgopal S, Dunnick J, Owusu-Ansah S, et al. Weather and Temporal factors associated with use of emergency medical services. *Prehospital Emerg Care*. 2019;23(6):802–10. <https://doi.org/10.1080/10903127.2019.1593563>. [published Online First: 2019/03/16].
39. Ramgopal S, Siripong N, Salcido DD et al. Weather and temporal models for emergency medical services: An assessment of generalizability. *Am J Emerg Med*. 2021;45:221–26. <https://doi.org/10.1016/j.ajem.2020.08.033> [published Online First: 2020/10/14].
40. Reddy VM, Judd O, Khalil H. Investigation of the influence of ambient temperature, atmospheric pressure and water vapour pressure on epistaxis admission rate. *Rhinology*. 2010;48(3):348–51. <https://doi.org/10.4193/Rhino10.020>. [published Online First: 2010/11/03].
41. Steul K, Kowall B, Oberndörfer D, et al. Rescue service deployment data as an indicator of heat morbidity in Frankfurt/Main, Germany (2014–2022)-Trend association with various heat exposure indicators and considerations for outreach. *Int J Hyg Environ Health*. 2023;254. <https://doi.org/10.1016/j.ijheh.2023.114250>.
42. Takada A, Kodera S, Suzuki K, et al. Estimation of the number of heat illness patients in eight metropolitan prefectures of Japan: correlation with ambient temperature and computed thermophysiological responses. *Front Public Health*. 2023;11. <https://doi.org/10.3389/fpubh.2023.1061135>.
43. Zhou ZY, Matteson DS, Woodard DB, Henderson SG, Micheas AC. A spatio-temporal point process model for ambulance demand. *J Am Stat Assoc*. 2015;110(509):6–15. <https://doi.org/10.1080/01621459.2014.941466>.
44. Kuo YH, Chan NB, Graham CA, et al. An integrated approach of machine learning and systems thinking for waiting time prediction in an emergency department. *Int J Med Informatics*. 2020;139:104143. <https://doi.org/10.1016/j.ijmedinf.2020.104143>.
45. Westgate BS, Woodard DB, Henderson SG, et al. Travel time Estimation for ambulances using bayesian data augmentation. *Annals Appl Stat*. 2013;7(2):1139–61. <https://doi.org/10.1214/13-AOAS626>.
46. Westgate BS, Woodard DB, Henderson SG, et al. Large-network travel time distribution Estimation for ambulances. *Eur J Oper Res*. 2016;252(1):322–33. <https://doi.org/10.1016/j.ejor.2016.01.004>.
47. Costs. Commitment and locality: a comparison of for-profit and not-for-profit health plans. *Inquiry: J Med Care Organ Provis Financing*. 2004;41(2):116–29. [https://doi.org/10.5034/inquiryjrn1\\_41.2.116](https://doi.org/10.5034/inquiryjrn1_41.2.116). [published Online First: 2004/09/29].
48. Cuschieri S. The STROBE guidelines. *Saudi J Anaesth*. 2019;13(Suppl 1):S31–4. [https://doi.org/10.4103/sja.SJA\\_543\\_18](https://doi.org/10.4103/sja.SJA_543_18). [published Online First: 2019/04/02].
49. Mechem CC, Yates CA, Rush MS, et al. Deployment of alternative response units in a High-Volume, urban EMS system. *Prehospital Emerg Care*. 2020;24(3):378–84. <https://doi.org/10.1080/10903127.2019.1657212>.
50. VAL-handboken-. Datalager för uppföljning av vårdhändelser i Region Stockholm [VAL-handbook-Datalayers and follow up in Region Stockholm]. In: Stockholm R, ed., 2023.
51. Ladda ner meteorologiska observationer <https://www.smhi.se>: The Swedish Meteorological and Hydrological Institute; 2023 [cited 2024-04-15]. Available from: <https://www.smhi.se/data/meteorologi/ladda-ner-meteorologiska-observationer#param=airtemperaturenstant,stations=core> accessed 2023/11/06 2023.
52. Griesbach C, Säfken B, Waldmann E. Gradient boosting for linear mixed models. *Int J Biostatistics*. 2021;17(2):317–29. <https://doi.org/10.1515/ijb-2020-0136>.
53. Schneider A, Hommel G, Blettner M. Linear regression analysis: part 14 of a series on evaluation of scientific publications. *Deutsches Ärzteblatt International*. 2010;107(44):776–82. <https://doi.org/10.3238/arztebl.2010.0776>.
54. Schober P, Vetter TR. Linear regression in medical research. *Anesth Analgesia*. 2021;132(1):108–09. <https://doi.org/10.1213/ANE.00000000000005206>.
55. Li M, Vanberkel P, Carter AJE. A review on ambulance offload delay literature. *Health Care Manag Sci*. 2019;22(4):658–75. <https://doi.org/10.1007/s10729-018-9450-x>.
56. Population per month by region, age and sex. Year 2000M01–2023M05 <https://www.statistikdatabasen.scb.se>: Statistics Sweden; 2023 [cited 2023-10-12]. Available from: [https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START\\_\\_BE\\_\\_BE0101\\_\\_BE0101A/BefolkManad/](https://www.statistikdatabasen.scb.se/pxweb/en/ssd/START__BE__BE0101__BE0101A/BefolkManad/) accessed 2023-08-17 2023.
57. Watanabe BL, Patterson GS, Kempema JM, et al. J Am Coll Emerg Physicians. 2019;74(1):101–09. <https://doi.org/10.1016/j.annemergmed.2018.09.032>. Is Use of Warning Lights and Sirens Associated With Increased Risk of Ambulance Crashes? A Contemporary Analysis Using National EMS Information System (NEMSIS) Data. *Annals of emergency medicine*.
58. Avtal Vågbyren ambulanssjukvårdstjänst AISAB. [Contract regarding EMS between Region Stockholm and AISAB]: Stockholms läns landsting, 2018.
59. DIREKTIV 8:2| Direktiv för passning [DIRECTIVE 8:2] Directive for fluid deployment| vardgivarguiden.se: Region Stockholm; 2020 [cited 2023-11-21]. Available from: <https://vardgivarguiden.se/globalassets/administration/patientadministration/prehospital-var/direktiv/passning-i-lanet.pdf> accessed 2023-11-21 2023.
60. Peyravi M, Khodakarim S, Örténwall P, et al. Does temporary location of ambulances (fluid deployment) affect response times and patient outcome? *Int J Emerg Med*. 2015;8(1):37. <https://doi.org/10.1186/s12245-015-0084-1>.
61. Breiman LJM. Random Forests. 2001;45:5–32.
62. Kriegeskorte N, Golan T. Neural network models and deep learning. *Curr Biol*. 2019;29(7):R231–36. <https://doi.org/10.1016/j.cub.2019.02.034>.
63. Wright R. Interpreting Black-Box Machine Learning Models Using Partial Dependence and Individual Conditional Expectation Plots. 2018.
64. Dzieciolowski K. Explaining black-boxmachine learning models. 2019.
65. Jerome HF. Greedy function approximation: A gradient boosting machine. *Annals Stat*. 2001;29(5):1189–232. <https://doi.org/10.1214/aos/1013203451>.
66. Swedish Ethical Review Authority. Approved Ethical review Dnr: 2022-03701-01. 2022.
67. Guideline for good clinical practice E6(R2.) <https://www.ema.europa.eu>: European Medicines Agency; 2016 [cited 2024-03-10]. Available from: <https://www.ema.europa.eu>.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.