BILINGUAL PUBLISHING GROUP

Journal of Electronic & Information Systems
https://journals.bilpubgroup.com/index.php/jeis

ARTICLE

The Application of Information Systems to Improve Ambulance Response Times in the UK

Alan Slater

University of Huddersfield, Huddersfield, HD13DH, UK

ABSTRACT

Emergency ambulance services in the UK are tasked with providing pre-hospital patient care and clinical services with a target response time between call connect to on-scene attendance. In 2017, NHS England introduced four new response time categories based on patient needs. The most challenging is to be on-scene for a life-threatening situation within seven minutes of the call being connected when such calls are random in terms of time and place throughout a large territory. Recent evidence indicates emergency ambulance services regularly fall short of achieving the target ambulance response times set by the National Health Service (NHS). To achieve these targets, they need to undertake transformational change and apply statistical, operations research and artificial intelligence techniques in the form of five separate modules covering demand forecasting, plus locate, allocate, dispatch, monitoring and re-deployment of resources. These modules should be linked in real-time employing a data warehouse to minimise computational data and generate accurate, meaningful and timely decisions ensuring patients receive an appropriate and timely response. A simulation covering a limited geographical area, time and operational data concluded that this form of integration of the five modules provides accurate and timely data upon which to make decisions that effectively improve ambulance response times.

Keywords: Ambulance response times; Demand forecasting; Geo-location models; Simulation

1. Introduction

Management in the UK Ambulance Service has accepted the challenge to develop the ability to collect and apply information innovatively. However, the focus has been on rapidly changing technology

*CORRESPONDING AUTHOR:
Alan Slater, University of Huddersfield, Huddersfield, HD13DH, UK; Email: alan@contactslater.co.uk

ARTICLE INFO
Received: 5 August 2023 | Revised: 28 August 2023 | Accepted: 29 August 2023 | Published Online: 21 September 2023
DOI: https://doi.org/10.30564/jeis.v5i2.5881

CITATION
Slater, A., 2023. The Application of Information Systems to Improve Ambulance Response Times in the UK. Journal of Electronic & Information Systems. 5(2): 10-24. DOI: https://doi.org/10.30564/jeis.v5i2.5881

and solving stand-alone issues such as crew scheduling rotas rather than linking opportunities together to improve overall efficiency and productivity. The centre of attention for management since the COVID-19 pandemic has been arresting a significant decline in ambulance response time (also known as 'clock-time'). This paper addresses the need to understand how identifying, upgrading and linking together several focused micro-information modules will create a macro information system that offers a foundation for improved ambulance response times.

The ambulance response time process is seen by management as the most significant issue in their service to critically ill patients because response time is both challenging and the key performance indicator by which the service is judged by both politicians and the public. This paper focuses upon an alternative approach to improving the ambulance dispatch process by utilising the combination of several real-time information systems which the ambulance service control directly on a day-to-day basis. There are, however, other influences on response times that cannot be controlled by the ambulance service on a day-to-day basis including vehicle design, vehicle maintenance, driver training, traffic management and public awareness. Attention to such issues may be combined with dispatch information to improve ambulance response times.

Attention to the information systems which support the ambulance response system could be the first step in the creation of a 'smart ambulance operation' which uses a set of rules incorporated in an 'artificial intelligence' system to recommend appropriate actions to call-handlers and dispatchers.

2. Operating environment

The objective of the ambulance service is to provide 'out-of-hospital' early medical assistance to 'save lives' by causing 'no further harm' before and during transporting a patient to a relevant hospital facility. Cooke [1] speculated that delays in pre-hospital care could lead to poorer patient outcomes and patient satisfaction increased when response was rapid.

In England, the public believes there is equality of access to appropriate pre-hospital care from the ambulance service based upon national ambulance response time targets—which are the time taken in minutes from an emergency call being connected to the ambulance service and the on-scene attendance by appropriate staff. The UK public has been encouraged by both the NHS management and the press to judge the ambulance service by its response to 'life-threatening' calls. Indicating that when they make a call for an ambulance one will be dispatched immediately (with 'blue lights and sirens') implying that resources are always available to answer every call and that each call is equally important and taken in priority of receipt.

In reality, ambulance managers prioritise calls using various response times according to patient needs. This means that calls classified as 'life-threatening' have a short response time target, whereas calls classified as 'urgent' have a longer response time target. This practice creates a tiered system of requirements to respond which allows resources to be allocated throughout the territory as 'ready to respond' thus maximising the response time target achievement with limited resources. However, short-term peaks in demand may lead to a lack of resources to respond to patient calls. In the event of a shortage of resources, an outbound crew on less urgent calls may be diverted to a life-threatening call.

The public may be unaware and do not recognise that there is a variation in response times although call handlers normally quote a forecast on-scene arrival time either when a despatcher has allocated a resource or alternatively they quoted a forecast waiting time before a resource will be allocated. The public is not made aware of the contributory factors behind any change in response time performance. Only since COVID-19 has such factors as traffic delays, hand-over delays at Accident and Emergency, or staff and vehicle availability been cited as causes for response time delays.

Unfortunately, recent press reporting concentrates upon failures to attend 'life-threatening emergencies' within the 7-minute response time target even though such situations represent less than 10% of all calls.

In reality, emergency ambulance services are tasked with reaching patients, by category of clinical need, as defined by the Advanced Medical Priority System (AMPDS), with suitably qualified staff. The response time targets, rather than the patient outcome, have become the key performance measure by which the press and public judge the success of ambulance services. However, Heath and Radcliffe [2,3] have criticised the NHS for concentrating only on ambulance service response times when the ambulance service offers a greater range of skills.

The NHS has over recent years reviewed the AMPDS and adjusted the provisional patient diagnosis categories, which are accessed by the call-handlers, such that the target response time is more appropriate to the patient pathway measured in terms of potential further harm and probability of a successful outcome. The NHS is currently developing response time targets for some specific conditions for the full clinical pathway from initial call to treatment in hospital and discharge to community care. These changes are based on data gathered by the NHS on patient outcomes from particular clinical pathways. Such applied information systems are available to the ambulance service to target and improve the management of ambulance response times.

When the NHS was first formed in 1947 the ambulance service was tasked with providing out-of-hospital care by attending on-scene emergencies providing patients with first-aid and transport to a suitable medical facility. In recent years the ambulance service provision has been expanded into other areas of NHS social care provision including situations that arise from mental health, alcohol or drugs and homelessness. These situations have arisen from a combination of changes in primary care strategy and shortages in the provision of social care. Privately, ambulance service managers indicate that many of their calls now relate to emergencies derived from a lack of social care where the ambulance service is the patient's last port of call.

This increase in workload placed pressure on the service by expanding demand, increasing on-scene time and increasing patient hand-over time at medical facilities. In addition, Gething, University of Wales, Health Board [4] reported a substantial increase in emergency calls mainly due to an expanding elderly population and the widespread misunderstanding amongst the general public of the ambulance service offering that all calls would obtain an immediate response.

The nature of a medical emergency indicates that demand for the ambulance service will be somewhat random in terms of 'time', 'place', and the 'required response'. The operational complexity of the ambulance response time problem is one of the most complex logistics tasks. The main difficulty is that in certain life-threatening situations there is an extremely short time before a patient may suffer further harm if no first-aid is administered. **Figure 1** outlines a process flow chart and the related urgency of responses required for certain conditions. With average demand for the busiest ambulance services at over one call per minute with demand emphasised by seasonal variations and event escalations the logistics requirements to achieve the response time targets are particularly difficult to manage. The challenge for operations managers is to recognise the urgency of the need, allocate appropriate resources, and achieve the target response time 24 hours/7 days a week/52 weeks a year at any location in the territory served whatever the traffic or weather conditions.

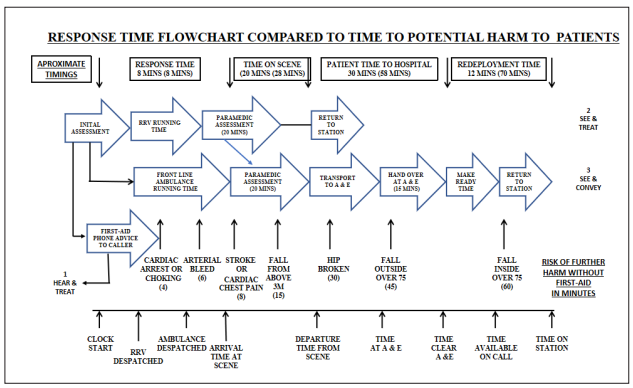

Figure 1. Response time flowchart compared to time to potential harm to patients.

3. Response time targets

Price [5] and Wankhade [6] indicated that ambu-

lance services could not keep pace with demand and were managing the 'response time' target by dispatching more than one resource to life-threatening calls rather than considering overall patient needs. Cooke [1] concluded that response time targets had been the main motivator for service restructuring although ambulance service management recognised that targets themselves do not save lives, more significantly ambulance availability, staff training and communications throughout the patient pathway would improve service and save lives.

Following a study into ambulance response times by Sheffield University, NHS England [7,8] established the Ambulance Response Programme with new integrated performance level and response time targets for all emergency calls based upon a combination of triage time by call handlers and ambulance response time by suitable crews for four categories of patient need:

Category 1—defined as 'life threatening' cases where the triage time is the earliest of 30 seconds from the call being connected, an ambulance being dispatched, or the patient needing to be identified. These cases require a 7-minute mean response with 90% of the calls responded to within 15 minutes or less. In addition, suitable transport must be available to convey the patient within 15 minutes of the call being connected.

Category 2—defined as 'emergency cases' where the triage time is the earliest of 240 seconds from the call being connected, an ambulance being dispatched, or the patient needing to be identified. These cases require an 18-minute mean response time with 90% of the calls responded to within 40 minutes or less. The 'clock' is only stopped by the arrival of a suitable vehicle to convey the patient, or if a vehicle is not required the first staff on the scene.

Category 3—defined as 'urgent cases' where the triage time is the earliest of 240 seconds from the call being connected or an ambulance being dispatched. These cases require a 90% response within 120 minutes by suitable transport, but if the patient does not require transport then the first staff on-scene stops the 'clock'.

Category 4—defined as 'non-urgent' cases where the triage time is 240 seconds from the call being connected, an ambulance being dispatched, or the patient needs to be identified. 90% of these cases should be responded to within 180 minutes and the 'clock' is only stopped by the arrival of a vehicle suitable to transport the patient, but if the patient does not require transport then the arrival of the first staff is on-scene.

Further integrated targets were also defined. For example, by 2022 90% of eligible heart attack patients should receive definitive treatment at a specialist heart centre within 150 minutes of the call being connected. These targets imply that ambulances take patients with certain defined conditions directly to the relevant hospital facility rather than Accident and Emergency (A & E).

4. Earlier theoretical solutions

In the past, operations research algorithms have offered a range of potential solutions addressing specific topics such as ambulance numbers and locations. Such studies have included Toregas et al. [9] who used optimisation to minimise the number of ambulances used to cover a defined geographical area. Whereas, Church and ReVelle [10] and Repede and Bernardo [11] indicated how fixed ambulance numbers could be located to obtain maximum geographical coverage. Gendreau et al. [12] and Dorner et al. [13] combined these two ideas to provide the minimum number of ambulances over a maximum geographical coverage. Daskin [14,15] modelled the impact of ambulance unavailability and Carter et al. [16] used queuing theory with fixed locations to tackle a similar problem. Larson [17] developed a 'hypercube' queuing model to select ambulances to respond from a fixed fleet. Lubicz and Meielczarek [18], Savas [19], Fitzsimmons [20], Swoveland et al. [21], Erkut et al. [22] and Inakawa et al. [23] have all used queuing and simulation techniques to predict ambulance response times in specific cities or geographical areas. Brotcorne et al. [24] provide a comprehensive review of ambulance location models. Similar specific parameters have been modelled by Fitch et al. [25], Blackwell

and Kaufman [26] and Shane et al. [27]. These operations research techniques were employed to tackle location and historic deployment issues rather than address the end-to-end tasks of managing a short lead time 24/7 emergency ambulance service supporting populations in a substantial territory

In addition, early operation research models were weak on input data and underlying operating assumptions and failed to provide credible solutions that represented an improvement upon existing operations. When these models were compared with real-time ambulance deployment planning there were several significant weaknesses. Furthermore, these models failed to consider significant real-time opportunities such as the re-deployment of returning ambulances to alternative locations, staggered crew shift times, diverting ambulances from less non-urgent tasks to life-threatening tasks, or the short-term use of alternative or third-party ambulance services for non-urgent cases.

Brotcorne et al. [24] indicated that several comments had been made by both users and reviewers of these techniques. Firstly, the data sets employed are historic, time-limited and do not account for a dynamic starting position, or peaks and troughs in demand on a seasonal, monthly, weekly, daily or hourly basis. Secondly, as shown by Carson and Batta [28], the travel time differences throughout the day are ignored and the travel time is either calculated on a straight-line distance at a constant speed or distance is calculated on the 'square root law' devised by Kolesar [29], leading to significant inaccuracies when compared with actual ambulance travel times. Thirdly, each model limited its scope by addressing only the significant issues of either location or fleet size, although Naoum-Sawaya and Elhedhli [30] considered a continuous-time chain to redeploy ambulances to a location upon completion of a task.

None of the models quoted coped with dynamic operational practices such as the use of deployment locations only on specific days or times of day; variable road speeds on different roads and the same road throughout the day. In addition, variations in on-site timings at different medical facilities including handover delays were not considered. Furthermore, they do not consider what has now become the operational practice in terms of dispatching more than one ambulance from different locations to a single call; the use of hubs (main ambulance stations) with satellites and despatch points (temporary standby ambulance locations), and the use of crews when returning from a task or extending crew shift times.

5. Operational issues

The shortfall in the operations research models indicates that problems facing the ambulance service are extremely complex and not those which could be solved by models which combine location selection with vehicle routing and scheduling because they involve a combination of several real-time operational issues including:

· Call pick-up delays at the call-centre when busy

· Insufficient or incorrect information to categorise patient needs correctly

· The immediate or subsequent availability of suitable crews

· Outbound (and inbound) traffic conditions

· Ability to obtain relevant patient information

· Selection or diversion to a suitable medical facility

· Waiting time and handover delays at hospital facilities

· Make ready and/or crew break before returning to a standby location

· Knowledge of current resources deployed on which category of task

· Progress of each deployment and potential crew reassignment time

· Currently available resources by type and skill level

· Number of patients on-scene at each location

· Data on the changing condition of the patient(s)

· Location of the incident and access to the patient(s)

· Time of day, traffic conditions, weather conditions

These issues delay the process flow of attending, on-site treatment and delivery of a patient to a suitable medical facility and may risk failing to achieve the target response time for the complete task. Lord Carter [31] suggested the reality of ambulance service

operations is currently very different from the optimum process due to the combination of increased demand and a need to substantially improve productivity. In addition to response times the ambulance service has a wider set of financial and operating key performance measures to report to NHS England monthly. However, in practice, there is real operational concentration on response times because both the general public and the national press highlight timing failures regularly.

Following receipt of an emergency call current practice is to follow one of three pathways—see **Figure 1** (where patient pathways are numbered 1, 2 and 3). Initially, the patient's need is categorised using AMPDS then call-centre staff will either deploy a suitable resource or provide telephone advice (known as 'hear and treat' 1). Some emergency and urgent calls may be concluded on scene employing treatment or advice from the crew (known as 'see and treat' 2). For life-threatening and emergency calls (known as 'see and convey' 3). one or more resources may be despatched immediately (if available) or existing outbound crews servicing urgent or less urgent calls may be diverted.

To assist with achieving target response times, call-centre staff have the electronic mapping of the territory which is combined with the use of full postcodes, or the compass on a caller's mobile phone, to give an accurate location of the patient. They have vehicle tracking and tracing to locate vehicles and coding to determine if vehicles are available and the category of the task they are undertaking. Workforce planning indicates the remaining shift time available to staff on duty and the potential availability of crews about to start a shift. On the local map, they have the locations of all medical facilities, on-call staff, community first responders, members of the hazardous area response team, the helicopter emergency medical service (HEMS), mountain rescue teams (MRT) and the location of static public access defibrillators and bleed control packs.

Call-centre staff also know about handover delays at each medical facility, vehicles out of service and crews on shift potentially taken out of service for operational reasons. On-scene staff need to know if they have multiple patients to consider at one site and the nature of any hazard that may impede staff either traveling to the site or at the site plus whether any other emergency service has been called out to assist and is a rendezvous arranged with anyone. Despatchers will monitor both the progress and requirement for those resources where they have despatched more than one resource to enable them to achieve the response time; so that they could stand down duplicated resources at the first opportunity.

In the event of severe life-threatening trauma being recognised by staff they may in addition despatch support in the form of a mobile critical care team consisting of a paramedic and/or doctor or the helicopter emergency medical service if suitable and available.

6. Alternative approach

The key issues to facilitate improving ambulance response times and staff productivity are the collection, analysis and availability of relevant timely information. To overcome the deficiencies encountered when using operations research techniques to optimise specific issues an alternative approach would be the combination of solutions from five real-time modules all of which could be integrated using a data-warehouse—see **Figure 2**.

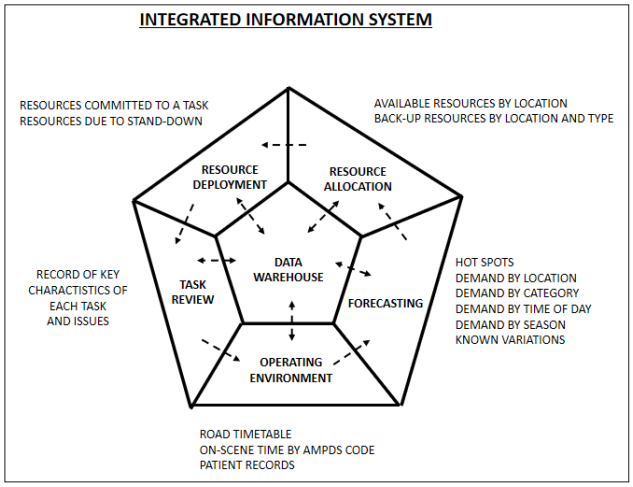

Figure 2. Integrated in formation system.

In this instance, a module is represented by one or more algorithms that obtain real-time or historic data from a directly linked data-warehouse to either

answer a specific question or add further processed data to the data warehouse.

6.1 Demand forecasting from historic data

Traditionally, the ambulance service has collected a large quantity of real-time data reflecting demand parameters including location of demand, timing of calls, classification of requirement, age of patients, operational conditions and short-term outcome for patients. This data indicates that calls are in essence random both in location and nature but analysis over an extended period of 3 years shows similar macro patterns if one-off factors such as mass-casualty incidents, pandemics, events or severe weather are discounted.

By combining ambulance service data with data from public services such as postcodes [32] and population data (see Office of National Statistics 2022) it is possible to determine some basic demand parameters. Pidd et al. [33] showed that mapping volume data by location and time using a geographic information system (GIS) would highlight demand 'hot spots'. Such analysis of both ambulance service and GIS data shows demand by postcode sector (defining area), showing potential forecasting variations, and defining demand by category of call, time of day and age of the population—see **Table 1**.

The profile may differ in each selected territory, however, data will be more accurate for Functional Urban Areas (FUA) which are densely inhabited areas in cities (urban and sub-urban) and slightly less populated commuter zones (semi-rural and rural). With data, collected over 3 years and weighted to the current situation by employing exponential smoothing, it is possible to plot forecast demand at postcode sector level and time of day. Best forecasts are obtained for categories 1 and 2 calls for a limited period (say 12 weeks) which account for seasonality and key conditions such as holiday peaks in demand. Using such demand data by day of the week and considering weather details, it is possible to forecast ambulance requirements to potential 'hot spots' by the time of day which allow staff to locate unallocated resources throughout the territory to areas with a high probability of demand and therefore maximise the opportunity to achieve the target response times.

6.2 Resource allocation

Previously resources were located at local ambulance stations and hospitals but with the development of direct-to-vehicle communications plus vehicle tracking and tracing the control centre has real-time knowledge of each crew's location (even while they are traveling). The current thinking is to allocate resources to ambulance stations and temporary deployment points in FUAs. Deployment points are parking places for a single front line double manned ambulance awaiting a job, often located at fire or police stations, shopping centres or garages where there are suitable 'comfort' facilities for the crew.

However, based upon significantly improved analysis of demand data it is now more appropriate to allocate resources based on forecast demand. 'Hot spots' could be geo-fenced providing a forecast of potentially successful travel time to a call based upon travel time from the centre point of a hot-spot. However, the target distance to potentially travel in a defined time from the centre of a 'hot-spot' is known as a 'geo-fence' which will vary by factors such as the resource available, the road layout and the forecast demand. by season, day of the week and hour of the day—see **Figure 3**. Ambulance stations and deployment points could be provided in areas of potential 'hot spots'. Significant gaps, not covered by 'geo-fences' could be filled by activating back-up resources including shift overlaps, on-call staff, Volunteer Ambulance Crews (VAC) or Community First Responders (CFRs). Once a crew is despatched it may be necessary to backfill the location with another crew at the earliest opportunity. This form of resource allocation improves the probability of achieving ambulance response times for category one and two calls.

Recognising that 'hot spots' alter throughout the day implies relocating operating and backup resources throughout the day as both demand patterns and travel times vary. Relocation is best undertaken at the beginning of a shift, as resources become free from a completed task or at the end of a crew break.

Table 1. Demand parameters: Data sheet.

GEOGRAPHIC AND POPULATION CHARACTERISTICS					
	Parameters	Post Code Split	Area	Population Daytime	Population Night
		%	%	%	%
URBAN	Below 0.5 square miles per post code	40	2	30	25
SUB-URBAN	Between 0.5 and 2.0 square miles per post code	30	8	36	40
SEMI-RURAL	Between 2.0 and 10.0 square miles per post code	20	20	26	25
RURAL	Over 10 square miles per post code	10	70	8	10
			100	100	100
		(Data averaged and rounded up)			

FORECASTABLE VARIATIONS IN DEMAND	
Season:	Spring - Summer - Autumn - Winter
Day of the Week:	Mon - Tues - Wed - Thu - Fri - Sat - Sun - Bank Holiday
Location Issues:	Tourist Holiday Season - School Holidays
Special Events:	Sporting Fixtures - County Shows - University Freshers Week
Zone:	Travel Time Dependant by Time of Day
Temperature:	Ice - Black Ice - Below Average - Above Average - Hot - Very Hot
Weather:	Snow - Sleet - Rain - Heavy Rain - Dry

FORECASTABLE LOCATION OF DEMAND
Urban - Sub-Urban - Semi-Rural - Rural
Population Demographics including Income Levels
Industry/Work Related - Ethnic Origin/Background Related - Housing/Social Related
Availability of Health and Social Services

CALL VOLUME BY TYPE AND TIME OF DAY				CALL VOLUME BY POPULATION AGE		
Time of Day (hours)	Life Threatening Emergencies	Emergency Calls		Age Group	Total Population Split	Demand by Age Group
	%	%			%	%
00.00 to 02.00	7	9		Under 10	12	9
02.00 to 04.00	5	6		11 to 20	13	5
04.00 to 06.00	2	3		21 to 30	12	7
06.00 to 08.00	7	6		31 to 40	15	5
08.00 to 10.00	9	7		41 to 50	14	8
10.00 to 12.00	11	9		51 to 60	13	10
12.00 to 14.00	11	10		61 to 70	10	14
14.00 to 16.00	10	10		71 to 80	7	18
16.00 to 18.00	10	10		81 to 90	3	16
18.00 to 20.00	10	11		Over 90	1	8
20.00 to 22.00	9	10			100	100
22.00 to 24.00	9	9				
	100	100				

Table 1 continued

CALL VOLUME BY TYPE AND TIME OF DAY				CALL VOLUME BY POPULATION AGE		
Time of Day (hours)	Life Threatening Emergencies	Emergency Calls		Age Group	Total Population Split	Demand by Age Group
UN-FORECASTABLE VARIATIONS IN DEMAND						
Severe Weather - Pandemic						
Mass Casualty Incident - Short Term Localised Issues						
Local Events						
Requirement for Mutual Aid						

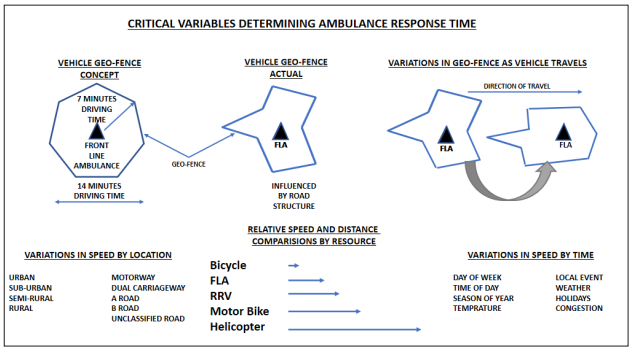

Figure 3. Critical variables determining ambulance response time.

6.3 Resource deployment

Traditionally the ambulance service has deployed the nearest available resource to a task. They have even redirected crews on-route to category 3 and 4 calls to a category 1 or 2 call if they were the nearest. However, if dispatchers adopt real-time vehicle tracking and tracing combined with geo-fencing methods they will be able to determine all potential options to attend the scene. Eglese et al. [34] indicated that the travel distance from any one point and each geo-fence structure is determined by the nature of the road speed and each vehicle geo-fence will be a different shape offering different ground coverage—see **Figure 3**. Eglese et al. [34] also showed that road speed differs between nodal points (road junctions on all except minor roads and tracks) on a specific road and that a 'road timetable' may be developed showing the estimated speed between each nodal point for each vehicle type with variations for the time of day, day of the week and season of the year. It is also practical to impose speed reductions for specific short-term events which are known to reduce vehicle travel speeds such as long-term road works.

With the use of vehicle tracking, geo-fencing based upon a relevant 'road timetable', it is possible to calculate the realistic area that may be covered in 7 minutes (for category 1 calls) or 18 minutes (for category 2 calls) from the vehicle's current location (whether static or mobile). Such an analysis of alternative vehicles and alternative routes may show the despatcher that a vehicle with the shortest travel distance could be slower than a vehicle with the shortest travel time.

It should be noted that urban and rural clusters will look different and there will be available resources traveling through each demand cluster when returning to a standby location and when outbound to a task—see **Figures 4 and 5**. The ambulance service could also use the combination of critical path programs, electronic mapping and vehicle tracking and tracing software to communicate in real-time with a caller who has a smart mobile the status and location of any outbound resource responding to their call—thus establishing a realistic expectation of an on-scene time and avoiding repeat requests for immediate ambulance services.

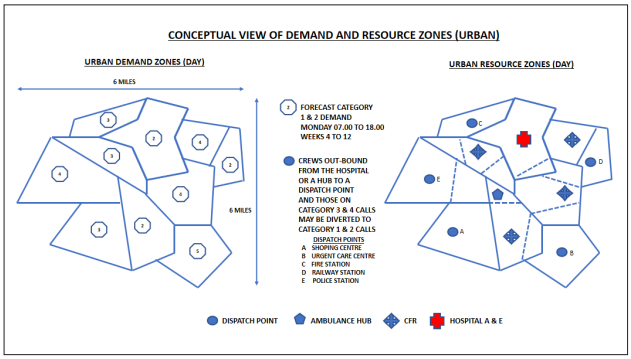

Figure 4. Conceptual view of demand and resource zones (Urban).

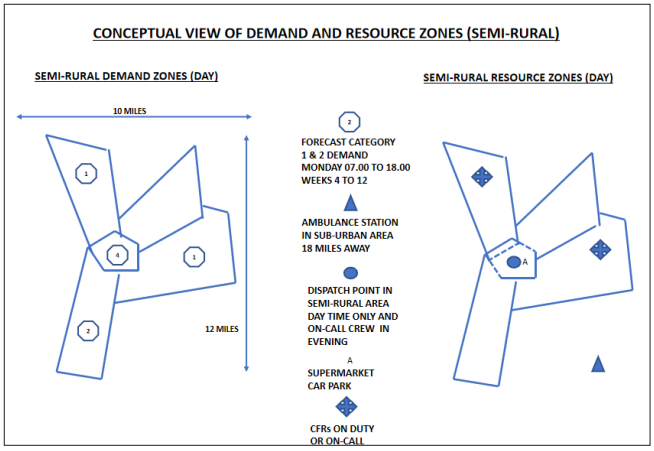

Figure 5. Conceptual view of demand and resource zones (Semi-rural).

6.4 Task review

One of the most significant benefits to a patient is the backup service provided by call-centre staff to both on-scene public before a crew arrives and the ambulance crew as they undertake their observations, conclude what actions are necessary, which pathway the patient should take and which medical facility the patient should be taken.

On-scene staff may also request assistance from senior clinicians within the call-centre or external supporting resources (for example Helicopter Emergency Medical Service—HEMS). In addition, staff may request the public on-scene to assist with immediate first-aid before an ambulance response arrives. This may take the form of advice by telephone, advice upon where to collect the nearest static defibrillator/bleed pack and how to use these resources.

Staff should be aware of all resources currently tasked, their location and when they are likely to become available (upon completion of a task or after any break) or when their shift finishes and whether it could be extended if necessary. Constant monitoring may indicate potential issues, for example, long ambulance handover times at a particular accident and emergency (A & E) department which may require further action to divert inbound ambulances to alternative accident and emergency departments or a local urgent care centre if practical.

6.5 Operating environment

To be able to improve the accuracy of vehicle running time and on-scene crew operating time any standard times for each activity must be amended to reflect the actual operating conditions at the time of the call. Such variations are achieved by monitoring, recording and updating specific operating tasks; for example:

The ‘road timetable’ is a database that forecasts the speed of a vehicle (operating with blue lights and sirens) between two nodal points at hourly intervals, by day of the week, under specific weather and traffic conditions during a specific season—an outbound travel forecast being the sum of the times between each nodal point on the route. The ‘road timetable’ is populated and updated from data gathered using the vehicle tracking and tracing system averaged to account for differed driver behaviours.

The ‘on scene operating time’ may be estimated for each AMPDS code, for each type and age of the patient, whether indoor or outdoor, at what time of day and in which geographic zone—collection of actual data over time could generate a ‘look-up’ table. Real-time monitoring against estimated will provide a clue as to whether the crew needs telephone or on-scene assistance.

Direct access (on a portable tablet) to patient records also provides the crew with a patient history, medication history and particularly which ‘primary’ and ‘secondary’ medical facilities they have visited. This may influence both questions asked during the crew’s initial observations and the decision upon an onward pathway to a medical facility. Previous patient calls to the ambulance service should be available on a database for crews to access while traveling outbound.

6.6 Data warehouse

A ‘data warehouse’ could be generated and updated from relevant data from each module 1 to 5 which allows relevant staff access 24/7/365. The data warehouse could be used to identify trends, recognise

resource deficiencies and highlight both existing and potential operating issues. Data could also represent the basis for comparisons between operations in different geographical zones supporting any necessary operational changes.

7. Simulation

To test the feasibility and practicality of developing and using the five modules a simulation was undertaken based on 24/7 operation, in a limited but representative geographical area, over the most difficult season of the year. A honeycomb pattern of adjusted hexagons was created based on a daily demand forecast at the response category level, corrected by exponential smoothing of historic demand patterns and broken down into six-hour periods. The honeycomb pattern covered an area with urban, suburban, semi-rural and rural populations. 'Hub' ambulance station locations were selected in functional urban areas, and 'deployment points' were selected in suitable facilities where there was up to 20 hours demand for 6 days each week. 'Spokes' (parking locations for Double Manned Front Line Ambulances DMFLA) were selected, throughout the rest of the territory, at other suitable facilities to cover forecast demand which was limited to specific periods in the day and days of the week.

Before the simulation was attempted a small data warehouse was established for the geographical area selected comprising three years of demand data, resources potentially available (based upon use), a deployment of resources at the starting position, data on local medical facilities, a high on-scene time by AMPDS code and a road time table (adjusted for 'blue-light' vehicles). Several basic decision rules were adopted, for example, road traffic accidents with two casualties would warrant two double manned front line vehicles (DMFLV). Additional demand was covered where known, for example, when the fire service dispatched three or more tenders to an incident one DMFLV would be dispatched to support the fire crews. No mass casualty incidents were included.

The simulation was undertaken based upon a geographically and time-limited data set broken into three years 'historic' data to produce a three-month statistical forecast of demand. Several operating rules were tested which achieved the required ambulance response times for the daily demand. These calculations were targeted to determine the maximum and minimum number and type of resources required at each hub, deployment point and spoke in the defined area. The simulation was used to determine how resources could be relocated throughout a shift pattern to meet forecast demand from the hub or deployment point to spokes and vice-versa to meet forecast demand within and between each hexagon.

The simulation was based on the defragmentation of the problem and split into four separate but interconnected principles:

Firstly, the forecast of short-term demand (expressed as potential calls by patient category and time of day) to determine the potential number of calls in each hexagon for every six hours in a rolling 24 hours. Calls were allocated an AMPDS code, day of the week and time of day based on weighted historic data. How many of these calls would fall into the 'hear and treat' and 'see and treat' groupings were estimated based on AMPDS coding. The AMPDS code would be translated into a category of patient need. A reduction factor was applied based on an estimate of how many would be duplicates or repeat calls for the same incident. Frequent caller data was identified but maintained in the dataset at category three because a response would be required. A separate forecast was calculated for calls transferred from the 111 service all of which were allocated to category two demand.

Secondly, based upon demand the selection of 'hubs', 'deployment points', 'spokes' and resources of all types would be initially allocated throughout the area. This would include a forecast of operational tasks by patient category in progress by location and percentage completion, so resources were limited at the start through a carry-over situation.

Thirdly, appropriate resources would be allocated to each call as it was forecast to occur (location, patient category by AMPDS code and time of day)—

the rules would be based upon achieving the target response time while maintaining maximum availability throughout the area. Resources that became available would be redeployed and their transit monitored.

Fourthly, all tasks would be monitored using a scheduling system with estimated task times and vehicle routing techniques to obtain operational data which would be used to determine when a task is likely to be completed and the crew available for the next task.

The simulation covered three months in the winter but excluded support for known local events, pandemics and any mass casualty occurrences. Each simulation was reviewed and run again with alternative operational rules.

The best results from several simulations indicated that 96% of the required ambulance response times would be achievable with 6% fewer resources than currently allocated in the area. However, this could only be achieved if and when resources became available they were immediately either allocated a break period or redeployed to areas of forecast demand. It was particularly important to ensure that 'hubs' would always have resources available or inbound following the completion of a task. The six most significant results include:

More than 15% of calls in FUAs were satisfied by crews being redeployed before reaching less urgent tasks or available in transit to a hub, deployment point or spoke.

Achievement for categories 1 and 2 patients was above the target emergency ambulance response time in FUAs but slightly below the target in other areas except where patients were initially attended by Community First Responders (CFRs) or on-call staff.

At periods of unusual peak demand times it would be necessary to convey category 4 patients using third-party ambulance services.

Over 20% of category 1 and 2 patients which achieved the target response time at peak times were responded to by crews from deployment points.

Less than 3% of category 1 and 2 calls required two or more crews

Spacing shift time starts, varying shift length and having a split shift for 'hub' based staff minimised the requirement to extend crew shifts to complete an allocated task.

The sensitivity of the simulation results showed that more than an 8% change in demand levels influenced whether the response time was achieved for both category 1 and category 2 even if such measures as extending crew shift time were employed.

After undertaking various sensitivity analysis runs, as stress tests based on the best simulation result several other significant conclusions arose, the six most significant were:

If hand-over times for category 2 and 3 patients at accident and emergency facilities doubled then significant local gaps appeared in resource availability at both hubs and deployment points leading to a need to convey category 4 patients employing third-party ambulance services.

The best results were obtained by having a paramedic as part of every crew because the paramedic would after observation treat certain AMPDS codes as 'see and treat'.

Utilising 'retained staff' in Rapid Response Vehicles (RRVs) from rural Primary Care surgeries and Volunteer Ambulance Crews (VAC) from rural fire, police or coastguard stations significantly improved the response time for category 1 and 2 patients in both semi-rural and rural areas.

Overall response time targets improved if crews were allocated category 3 and 4 patients at the end of their shift and shift starters were deployed to 'deployment points' and 'spokes' where they would initially be tasked with responding to category 1 and 2 patients.

Redirecting patients to either specialist medical facilities or alternative accident and emergency units (rather than the nearest one if hand-over times there exceeded the target) had little impact on the redeployment of crews or overall achievement of the target ambulance response times except where a patient could be grouped as 'hear and treat' and directed by call-centre staff to attend alternative medical facili-

ties or a pharmacy.

Following Hamet and Tremblay [35] the application of operational rules targeted at testing an Artificial Intelligence (AI) response towards testing call-centre staff efficiency by limiting resource options to those which could either achieve the target response time or were at the nearest 'hub'. These limited computing calculation time and offered the dispatcher a choice of the best results.

8. Conclusions

It has been widely understood that ambulance services in England have through their Computer Aided Dispatch (CAD), AMPDS and patient record systems collected a large amount of patient and operations data. However, updating, validating and processing this information in real-time has not been recognised as the backbone to achieving national emergency ambulance response times.

To address this problem the ambulance service could defragment the problem and using several currently available information technologies consider five key decisions in real time.

Where are the demand clusters for life-threatening and emergency calls?

How many of each type of resource are required and where to locate these resources?

How many, of which resources and when to allocate resources to each call?

To which facility (if further treatment is required) the patient should be conveyed?

Which location an ambulance should be deployed to upon completion of the allocated task?

It has been recognised that these questions may be answered by separate modules in the form of:

GIS analysis and forecasting modules to determine demand 'hot spots' and demand clusters by location and time.

A resource coverage and availability module matched to statistical forecasting of demand over a short-term time, based upon hexagons each with a radius set by travel distance over the response time requirements.

A recommended call resource allocation module based upon interpreting both a set of basic rules and past dispatch behaviour to maintain maximum availability of resources covering the territory.

A task scheduling module to provide the basis for a resource allocation and redeployment module based upon matching forecast short-term demand with shift time remaining for each operating crew and alternative resources that may be available.

A queuing theory module that monitors availability at each secondary medical facility (also defining specialisations) servicing the territory to deliver the patient to the most appropriate facility, minimise patient queuing and maximise crew turn-round.

Simulation and sensitivity analysis have proved that linking the technologies in these modules together, by utilising a data warehouse in real-time, provides an opportunity to understand short-term demand and be able to resource enough calls within the national emergency ambulance service target response times.

These developments represent the first step towards 'smart ambulance operations' by establishing the groundwork upon which information systems may be used in the decision-making process Subsequently, artificial intelligence will use a combination of activities to locate 'stand-by' crews, dispatch crews or alternative immediate assistance, select a receiving hospital, sending on-scene real-time internet video recordings to medical teams and to communicate patient observations directly from equipment in the ambulance to the hospital based medical team. Such activities are targeted to improve the medical assistance available throughout the patients' pathway and thereby improve the patient experience and outcome.

Potential progress in the application of these techniques nationwide may be limited by the ability to allocate sufficient financial support and the ability to attract staff with relevant information technology skills.

Author Contributions

This article has been researched and written by a single author.

Conflict of Interest

There is no conflict of interest.

References

[1] Cook, M., 2011. An introduction to the new ambulance clinical quality indicators. Ambulance Today. 7(5), 35-36.

[2] Heath, G., Radcliffe, J., 2007. Performance measurement and the English ambulance service. Public Money and Management. 27(3), 223-228.

[3] Heath, G., Radcliffe, J., 2010. Exploring the utility of current performance measures for changing roles and practices of ambulance paramedics. Public Money & Management. 30(3), 151-158.

[4] Gething,V (2015) .Written Statement – Clinical Review of Ambulance Response Time Targets' Cabinet Statement, Welsh Government, Health Board. https://www.gov.wales/written-statement-clinical-review-ambulance-response-time-targets

[5] Price, L., 2006. Treating the clock and not the patient: Ambulance response times and risk. BMJ Quality & Safety. 15(2), 127-130.

[6] Wankhade, P., 2011. Performance measurement and the UK emergency ambulance service: Unintended consequences of the ambulance response time targets. International Journal of Public Sector Management. 24(5), 384-402.

[7] Turner J (2017) 'How we remodelled Ambulance Services in England' Research Features, University of Sheffield, England. https://www.sheffield.ac.uk/research/features/remodelled-ambulance-service

[8] On the Day Briefing: Ambulance Response Time Programme [Internet]. NHS Providers. Available from: https://nhsproviders.org/resources/briefings/on-the-day-briefing-ambulance-response-programme#

[9] Toregas, C., Swain, R., ReVelle, C., et al., 1971. The location of emergency service facilities. Operations Research. 19(6), 1363-1373.

[10] Church, R., ReVelle, C., 1974. The maximal covering location problem. Regional Science. 32(1), 101-118.

[11] Repede, J.F., Bernardo, J.J., 1994. Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky. European Journal of Operational Research. 75(3), 567-581.

[12] Gendreau, M., Laporte, G., Semet, F., 1997. Solving an ambulance location model by tabu search. Location Science. 5(2), 75-88.

[13] Doerner, K.F., Gutjahr, W.J., Hartl, R.F., et al., 2005. Heuristic solution of an extended double-coverage ambulance location problem for Austria. Central European Journal of Operations Research. 13(4), 325-340.

[14] Daskin, M.S., 1983. A maximum expected covering location model: Formulation, properties and heuristic solution. Transportation Science. 17(1), 48-70.

[15] Daskin, M., 1987. Location, Dispatching and Routing Model for Emergency Services with Stochastic Travel Times [Internet]. Available from: https://www.semanticscholar.org/paper/LOCATION%2C-DISPATCHING-AND-ROUTING-MODELS-FOR-WITH-Daskin/5bc91cb1642f023fc1958eaf31c228b291f4cde4

[16] Carter, G.M., Chaiken, J.M., Ignall, E., 1972. Response areas for two emergency units. Operations Research. 20(3), 571-594.

[17] Larson, R.C., 1974. A hypercube queuing model for facility location and redistricting in urban emergency services. Computers & Operations Research. 1(1), 67-95.

[18] Lubicz, M., Mielczarek, B., 1987. Simulation modelling of emergency medical services. European Journal of Operational Research. 29(2), 178-185.

[19] Savas, E.S., 1969. Simulation and cost-effectiveness analysis of New York's emergency ambulance service. Management Science. 15(12), B-608.

[20] Fitzsimmons, J.A., 1973. A methodology for

emergency ambulance deployment. Management Science. 19(6), 627-636.

[21] Swoveland, C., Uyeno, D., Vertinsky, I., et al., 1973. Ambulance location: A probabilistic enumeration approach. Management Science. 20, 686-698.

[22] Erkut, E., Ingolfsson, A., Budge, S., 2007. Maximum Availability Models for Selecting Ambulance Station and Vehicle Locations: A Critique [Internet]. Available from: https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=965f00e8ec5caa3d08ab916b6c71528f5908a311

[23] Inakawa, K., Furuta, T., Suzuki, A. (editors), 2010. Effect of ambulance station locations and number of ambulances to the quality of the emergency service. The Ninth International Symposium on Operations Research and Its Applications (ISORA'10); 2010 Aug 19-23; Chengdu, China. p. 340-347.

[24] Brotcorne, L., Laporte, G., Semet, F., 2003. Ambulance location and relocation models. European Journal of Operational Research. 147(3), 451-463.

[25] Fitch, J., 2005. Response times: Myths, measurement and management. Journal of Emergency Medical Services. 30, 46-56.

[26] Blackwell, T.H., Kaufman, J.S., 2002. Response time effectiveness: Comparison of response time and survival in an urban emergency medical services system. Academic Emergency Medicine. 9(4), 288-295.

[27] Henderson, S.G., Mason, A.J., 2004. Ambulance service planning: Simulation and data visualisation. Operations research and health care: A handbook of methods and applications. Springer: Berlin. pp. 77-102.

[28] Carson, Y.M., Batta, R., 1990. Locating an ambulance on the Amherst campus of the State University of New York at Buffalo. Interfaces. 20(5), 43-49.

[29] Kolesar, P., 1975. A model for predicting average fire engine travel times. Operations Research. 23(4), 603-613.

[30] Naoum-Sawaya, J., Elhedhli, S., 2013. A stochastic optimization model for real-time ambulance redeployment. Computers & Operations Research. 40(8), 1972-1978.

[31] Carter, L., 2018. Operational Productivity and Performance in English Ambulance Trusts: Unwanted Variations [Internet]. Available from: https://www.england.nhs.uk/wp-content/uploads/2019/09/Operational_productivity_and_performance_NHS_Ambulance_Trusts_final.pdf

[32] Royal Mail (2015) 'History of UK Postcodes' Royal Mail Group, London. https://www.poweredbypaf.com/the-history-of-uk-postcodes/

[33] Pidd, M., De Silva, F.N., Eglese, R.W., 1996. A simulation model for emergency evacuation. European Journal of Operational Research. 90(3), 413-419.

[34] Eglese, R., Maden, W., Slater, A., 2006. A road timetableTM to aid vehicle routing and scheduling. Computers & Operations Research. 33(12), 3508-3519.

[35] Hamet, P., Tremblay, J., 2017. Artificial intelligence in medicine. Metabolism. 69, S36-S40.