HEALTH SYSTEMS
2022, VOL. 11, NO. 3, 161–171
https://doi.org/10.1080/20476965.2021.1908176

ORIGINAL ARTICLE

OPEN ACCESS

Using a genetic algorithm to solve a non-linear location allocation problem for specialised children's ambulances in England and Wales

Enoch Kung[a], Sarah E. Seaton[b], Padmanabhan Ramnarayan[c] and Christina Pagel[a1,2]

[a]Department of Mathematics, UCL Clinical Operational Research Unit, University College London, London, UK; [b]Department of Health Sciences, University of Leicester, Leicester, UK; [c]Children's Acute Transport Service (CATS), London, UK

ABSTRACT
Since 1997, special paediatric intensive care retrieval teams (PICRTs) based in 11 locations across England and Wales have been used to transport sick children from district general hospitals to one of 24 paediatric intensive care units. We develop a location allocation optimisation framework to help inform decisions on the optimal number of locations for each PICRT, where those locations should be, which local hospital each location serves and how many teams should station each location. Our framework allows for stochastic journey times, differential weights for each journey leg and incorporates queuing theory by considering the time spent waiting for a PICRT to become available. We examine the average waiting time and the average time to bedside under different number of operational PICRT stations, different number of teams per station and different levels of demand. We show that consolidating the teams into fewer stations for higher availability leads to better performance.

ARTICLE HISTORY
Received 24 August 2020
Accepted 1 February 2021

KEYWORDS
OR in health services
location-allocation analysis
emergency medical service
integer programming
genetic algorithms

1. Introduction

Paediatric intensive care services have been centralised in England and Wales since 1997. Children in need of intensive care are transported from their local district general hospitals (DGHs) to one of 24 paediatric intensive care units (PICUs) by specialist paediatric intensive care retrieval teams (PICRTs). The specialist teams provide patients with intensive care services as soon as possible. Ramnarayan et al. showed that using specialist teams to transport critically ill children improved survival to PICU discharge (Parslow et al., 2010). The current national quality standard is for PICRTs to reach the child's bedside at the DGH within 3 hours of accepting the referral (Paediatric Intensive Care Society, 2015). There are currently 11 PICRT locations across England and Wales, each with 1–2 teams available at any one time. A typical retrieval can take 4 to 5 hours, during which time the PICRT team is unavailable for additional referrals. When there is no team available for a referral (if all teams are out on a retrieval), a sick child at the DGH will have to wait longer for specialist care. While relatively rare, such delays do happen, particularly in winter when demand for PICRT services almost doubles (Ramnarayan S. et al., 2018) (Ramnarayan S. et al., 2015). With such a centralised service, important questions are: how many PICRT locations should there be? How many teams should staff each location? And where should PICRTs be physically located to best serve the population, where "best" is defined by stakeholders, but could include minimising average time to bedside, minimising the probability that any child waits longer than 3 hours for a PICRT arrival or minimising time for the child to arrive at the receiving PICU.

There is currently national variation among the different PICRT locations in how quickly they reach the patient's bedside and in the types of interventions routinely performed by each team during the transport episode (Parslow et al., 2010) (Paediatric Intensive Care Audit Network, 2020). The DEPICT study (Parslow et al., 2010) is a national, mixed methods study, the aim of which is to assess the impact of these variations on clinical outcomes and the experience of stakeholders (patients, families and healthcare staff). One strand of the DEPICT work is to use location allocation modelling to explore how changing the location of current PICRTs, the number of locations of PICRTs and the number of teams at each location, impacts on the time to bedside and other service metrics. Time to bedside could be reduced, for instance, by changing the locations and the number

CONTACT Christina Pagel c.pagel@ucl.ac.uk UCL Clinical Operational Research Unit, University College London, London, UK.
[1]The DEPICT Study Group members are: Elizabeth Draper, Paul Mouncey, Jo Wray, Matthew Entwistle, Anna Pearce, Fatemah Rajah, Will Marriage, Patrick Davies, Eithne Polke, Rachel Lundy, Ruth Evans, Emma Hudson, Victoria Barber, Steve Morris, Laura Drikite.
[2]The DEPICT project is funded by the National Institute for Health Research Health Services and Delivery Research programme (Project No:15/136/45). The views and opinions expressed therein are those of the authors and do not necessarily reflect those of the NIHR HS&DR programme or the Department of Health.

of PICRT locations to reduce journey times or by increasing the number of teams at each location to reduce the chance of a team not being available for a new referral.

In previous work (King et al., 2019), we used a simple p-median model to explore the achievability of the 3 hour service standard with the existing provision and location of PICRTs and whether there was room for improving time to bedside with a different allocation of PICRTs. In that initial work, we assumed that a retrieval team would always be available for a referral and that journey times and demand for services were constant. Under those assumptions, we showed that 98% of retrieval demand can be met within the 3-hour standard. Furthermore, we showed that the reduction from 11 locations to 8 locations lowers the percentage of demand covered in 3 hours only marginally. On the other hand, if the recommended time to bedside was made more stringent, fewer DGHs would be accessible within the standard, even with more PICRT locations.

In this work, we significantly extend the previous formulation by developing an optimisation framework that generalises the model by dropping some of the unrealistic assumptions in (King et al, 2019), in particular the assumption of constant travel times and the assumption that a team will always be available for retrieval. This paper is organised as follows: a brief literature review of how location allocation methods have been used previously to inform health services decision-making; in Section 2, a detailed description of the optimisation framework that allows stochastic journey times and incorporates queuing theory; in Section 3, a description of the computational methods that implement this non-linear optimisation framework; in Section 4, examples based on real data showing the method in practice; a discussion of the framework and next steps.

2. Literature review

Location allocation analysis is a field with many and diverse applications. The placement of facilities such as factories, schools, and even emergency services have been investigated in detail. Location allocation analysis has also, building upon basic models, progressed from a static and deterministic approach to incorporate the dynamics and stochastic nature that more closely resembles real situations. The (Toregas et al., 1971) and Li (Li et al., 2011) provide an overview of the basics of location allocation analysis and its historical development. In this section, we discuss examples from the academic literature that are particularly relevant to our work.

A general objective in location allocation research is to place facilities in a network with a view to maximise the effectiveness of the network. The popular metrics for effectiveness could be to reduce the total necessary travel distance or time to facilities, such as those used for *p*-median problems (Toregas et al., 1971)(Zhao et al., 2011). Another metric is to maximise the coverage of the population with the minimal number of stations, which is used in the location set and maximal coverage location problems. Problem-specific metrics can be used depending on the area of application. For example, ambulance location problems can use expected patient survival as a measurement of the quality of facility locations, such as the Maximal Survival Location Problems in (Ingolfsson et al, 2009) (Coates & McCormack, 2015). Most problems are solved numerically with optimisation packages or simulation methods.

A variety of extensions have been developed for location allocation analysis. Stochastic elements have been introduced in many ways, including the use of queuing theory and probability. The work of Mirchandani (Mirchandani & Odoni, 1979) considered the fluctuations of travel time due to traffic conditions. In a paper by Daskin (Daskin, 2008), a covering model is developed to account for the possible unavailability of facilities. The hypercube queueing model by Larson (Berman et al., 1987) was employed to model the state of service availability of facilities as a more convenient setting to search for optimal location placements. Church and Revelle (Church & ReVelle, 1989) introduced stochasticity into the covering model by guaranteeing coverage to those demanding service with a likelihood above a certain threshold probability which was later extended to also consider service availability.

Location analysis has been applied to the realm of police, ambulance, and other emergency services. Larson (Larson, 1974) used the hypercube queuing model to divide a district into police patrol beats so as to equalise the workload of each police patrol while minimising response time. The ambulance network was similarly studied (Ingolfsson et al., 2009)(Shiah and Chen, 2007) with a goal to minimise the time of arrival to incident with extensions to include parameters such as capacity requirements and ambulance availability (PKnight et al, 2012). Queuing models such as the Priority Queuing Covering Location Problem (Silva & Serra, 2016) have been applied to emergency services as a covering model that allows prioritisation of calls for service. Further analyses have been done on emergency service planning such as the trade-off between equity and efficiency in the distribution of service to urban and rural areas (Burkey et al., 2012) (Dale, 1979).

In this paper, we use a stochastic *p*-median approach to model the paediatric intensive care retrieval system where queuing for the service is considered, expanding upon the approach of Berman (Berman et al., 1987) by solving both the optimal locations of

PICRT stations as well as the optimal allocation of DGHs to these stations.

3. Method

The basic elements of the objective function are the three different elements of a PICRT's round trip during a retrieval, illustrated in Figure 1. The three elements are (i) the journey from PICRT station to the District General Hospital (DGH) (green), (ii) the journey from the DGH to the receiving Paediatric Intensive Care Unit (PICU) (purple), and (iii) the return trip from the PICU to the base PICRT station (orange). Extending our previous approach of only considering the time to bedside (P. Ramnarayan et al., 2019), we now allow the other journey elements to contribute to the objective function, with their contribution weighted according to their importance as determined by the service user. Another key reason for including all legs of the journey is that total time away from base determines the availability of a team for a new referral. Additionally, a journey includes a mobilisation time for PICRTs between referral and leaving the PICRT base, the time it takes for a team to get to the child's bedside after a PICRT arrives at the DGH, and a period of treatment at the DGH before transporting the patient to the PICU; these times are assumed to be constant and set at $T_m = 30$ minutes, $T_a = 10$ minutes, and $T_p = 2$ hours, respectively, where these estimates are based on historical PICRT audit data.

Journey times between hospitals are scaled by demand, where demand is the number of requests for PICRT services from each DGH, so that journeys taken more frequently are given more weight. While demand can vary throughout the year or even time of day, demand is assumed to be constant within our optimisation framework. The impact of different demand levels is instead examined by exploring different realisations of the optimisation model under different scenarios of demand (e.g., winter vs non-winter).

After introducing our mathematical notation, we first develop the optimisation framework assuming that a PICRT is always available before extending this formulation to allow for a waiting time before a team becomes available for retrieval.

3.1. Interim formulation assuming that a team is always available

We first introduce some notation (see also Table 1). Let $\mathcal{I}$, $\mathcal{J}$, and $\mathcal{R}$ denote the set of DGHs, PICRT stations, and PICUs, respectively. The size of these sets are correspondingly denoted by I, J, and R. The set $\mathcal{J}$ comprises of both the PICUs in the set $\mathcal{R}$ and other existing PICRT station locations. Of the J PICRT stations, we limit the number of operating stations to be N. For example, in previous work (King et al., 2019), $N = 11$.

Decision variables that identify the operational stations and the hospital allocations are given by

$$X_j = \begin{cases} 1 & \text{station j is operational} \\ 0 & \text{otherwise} \end{cases} \qquad Y_{ji} = \begin{cases} 1 & \text{if station j serves hospital i} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

$$\text{for } i = 1, \ldots, I \text{ and } j = 1, \ldots, J.$$

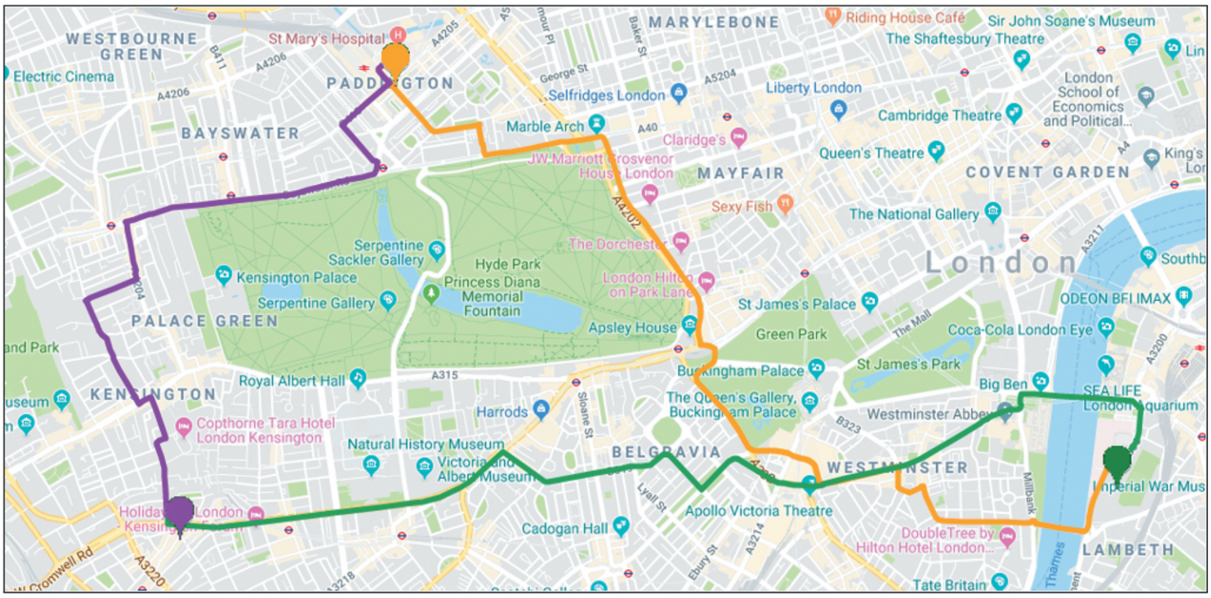


Figure 1. Illustration of a PICRT team's ROUND TRip: PICRT station to the District General Hospital (DGH) (green), DGH to the receiving Paediatric Intensive Care Unit (PICU) (purple), PICU back to the base station (orange).

Table 1. Notations.

X_j	1 if station j is operational, 0 otherwise
Y_{ji}	1 if hospital i is served by station j, 0 otherwise
$\mathcal{I}, I$	set of DGHs, number of DGHs
$\mathcal{J}, J$	set of potential PICRT stations, number of potential PICRT stations
$\mathcal{R}, R$	set of PICUs, number of PICUs
N	number of operational PICRT stations
d_i	demand of PICRT services over a year for hospital i
w_k	weight of the k-th journey for $k = 1, 2, 3$
$r(i)$	$r(i)$ is the closest PICU to hospital i
t^1_{ji}	journey time from PICRT station j to hospital i
$t^2_{ir(i)}$	journey time from hospital i to PICU $r(i)$
$t^3_{r(i)j}$	journey time from PICU $r(i)$ back to PICRT station j
T_m	mobilisation time; $T_m = 30$ minutes is assumed
T_a	time from arrival to hospital to patient bedside; $T_a = 10$ minutes is assumed
T_p	treatment period before transport to PICU; $T_p = 2$ hours is assumed
T_{ji}	the combined travel time $T_m + t^1_{ji} + T_a + T_p + t^2_{ir(i)} + t^3_{r(i)j}$
Z_i	$Z_i = j$ if hospital i is served by station j ($Y_{ji} = 1$)
λ_j	request of service per minute for station j
μ_j	service rate: number of customers served per minute for station j
c_j	number of retrieval teams at station j
ρ_j	utilisation rate: $\frac{\lambda_j}{c_j \mu_j}$

The objective function is a weighted sum of the three journey times in a PICRT's round trip. For the trip between the DGH and PICU, we calculate the expected travel time averaged proportionately over all possible PICU destinations from each DGH. The travel times of a team's journey are labelled t^1_{ji} (station j to DGH i), $t^2_{ir(i)}$ (DGH i to PICU $r(i)$), and $t^3_{r(i)j}$ (PICU $r(i)$ to station j). These times are scaled by the demand of each hospital, denoted by d_i, and also by weighting the different parts of the journey by a parameter w_k, where $k = 1, 2, 3$.

The result is a linear integer optimisation problem

$$\min_{(\{X_j\},\{Y_{ji}\})} \sum\nolimits_{i,j} d_i Y_{ji} (w_1 (T_m + t^1_{ji} + T_a) + w_2 (T_p + t^2_{ir(i)}) + w_3 t^3_{r(i)j}) \quad (2)$$

subject to the following constraints:

(C1) $\sum\nolimits_j Y_{ji} = 1$ for $i = 1, \ldots, I$ (each hospital is served by one and only one PICRT station)

(C2) $Y_{ji} - X_j \leq 0$ for $i = 1, \ldots, I$ and $j = 1, \ldots, J$ (hospitals must be served by an operational station)

(C3) $\sum\nolimits_j X_j \leq N$ (the number of operational stations are at most a fixed number N.)

(C4) $X_j, Y_{ji} \in \{0, 1\}$ (the variables are binary),

which can be solved by standard integer optimisation packages, using mean travel times between pairs of hospital (either estimated from historical data or using online software such as Google Maps).

3.2. Extension of problem to relax assumption that a team is always available

We now consider the possibility that a patient might need to wait for a team to become available to receive the next request. Incorporating a waiting time term introduces nonlinearity to the objective function in equation 2, requiring a new method for its solution. In this section, we detail how the waiting time is approximated using queueing theory and how a genetic algorithm is applied to solve the resulting problem.

3.3. Incorporating waiting time within our objective function

Each team at the PICRT station can be considered as a server and the patients as forming a queue, waiting for the server to be free. The queue can be arbitrarily long as patients are not in a physical queue but are waiting in local hospitals. We also assume that the service is first-come-first-serve. As is standard in modelling demand for emergency services, arrivals are taken to be memoryless (i.e., following a Poisson distribution). We cannot however assume such simplicity for the service time (the time that a retrieval team is away from base). Therefore, in PICRT station j, we have an $M/G/c_j/\infty$ queue. The parameter c_j refers to the number of teams working in PICRT station j. The waiting time can be calculated by simulation, which in our particular problem is computationally expensive, or approximated by an explicit formula.

A result from queueing theory allows us to approximate waiting time at steady state starting from a simpler $M/M/c_j/\infty$ queue, where the waiting time can be explicitly stated in the following way: for station j, we need the parameters

1. total arrival rate: $\lambda_j = \sum\nolimits_i d_i Y_{ji}$
2. service rate: $\mu_j = \frac{\lambda_j}{\sum\nolimits_i d_i Y_{ji} \mathbb{E}[T_{ji}]}$
3. utilisation rate: $\rho_j = \frac{\lambda_j}{c_j \mu_j}$,

where T_{ji} is the random variable of the round-trip travel time of a team from leaving the station to pick up a patient and returning to the station. Recall from the previous section that T_{ji} is dependent on which PICU a patient is transported to. Note also that the service rate is of this form because it is the inverse of the average service time at station j, which is the average of round-trip journey times over all hospitals i allocated to the station weighted by the demand of each of these hospitals.

These parameters allow us to approximate the waiting time of a $M/G/c_j$ queue at a stationary state using Kingman's formula (Kingman J.F.C., 1961), stated below (equation (3)), which states that the steady state waiting time of $M/G/c_j$ queue is the steady state waiting time of a $M/M/c_j$ queue (Allen, 1990) multiplied by a factor calculated from the coefficient of variation of the service time $C_{s,j}$:

$$W_j \approx \frac{(1 + C^2_{s,j})}{2} [\text{waiting time of } M/M/c_j/\infty \text{ queue at stationary state}] \quad (3)$$

$$\approx \frac{(1+C_{s,j}^2)}{2}\left[\sum{}_{\cdot k=0}^{c_j-1}\frac{(c_j\rho_j)^k}{k!}+\frac{(c_j\rho_j)^{c_j}}{c_j!}\frac{1}{1-\rho_j}\right]^{-1}\cdot \frac{\rho_j(c_j\rho_j)^{c_j}}{c_j!(1-\rho_j)^2\lambda_j}$$

The mean $m_{s,j}$ is the average service time and $\sigma_{s,j}$ the standard deviation; these can be estimated from the distribution of the travel times obtained from historical audit data (data on all PICRT journeys has been collected nationally since 2012). Given $\mathbb{E}(T_{ji})$ and $Var(T_{ji})$, we can calculate these two parameters by

$$m_{s,j} = \frac{1}{\mu_j} = \sum{}_{\cdot i}\frac{d_i}{\lambda_j}Y_{ji}\mathbb{E}[T_{ji}]$$

$$\sigma_{s,j}^2 = \mathbb{E}\left[\left(\sum{}_{\cdot i}\frac{d_i}{\lambda_j}Y_{ji}T_{ji}\right)^2\right] - \left(\sum{}_{\cdot i}\frac{d_i}{\lambda_j}Y_{ji}\mathbb{E}[T_{ji}]\right)^2$$

$$= \sum{}_{\cdot i}\left(\frac{d_i}{\lambda_j}\right)^2 Y_{ji}Var(T_{ji}).$$

Finally, our objective function including waiting time is given by:

$$\min_{(\{X_j\},\{Y_{ji}\})} F(\{X_j\},\{Y_{ji}\}) = \sum{}_{\cdot i,j}d_iY_{ji}(W_j + w_1t_{ji}^1 + w_2t_{ir(i)}^2 + w_3t_{r(i)j}^3), \tag{4}$$

$$\text{subject to the constraints (C1), (C2), (C3), and (C4).}$$

This objective function is no longer linear and requires a different approach to solving it in a reasonable computational timeframe.

3.4. Solving the non-linear problem using a Genetic Algorithm

We apply a genetic algorithm to approach this non-linear optimisation problem. While there are various possible heuristic approaches to solve this optimisation, we chose to use a genetic algorithm because of its simplicity in sorting through a vast pool of potential solutions and ability to combine fragments of optimal features from the population of solutions. Another advantage is that it can be implemented simply and is flexible enough to handle different objective functions for future use. For a detailed introduction to genetic algorithms, see (Mitchell, 1990). The main challenges in its application are the number of variables and the inclusion of constraints. We overcome both obstacles by restructuring our optimisation problem.

Instead of solving both the optimal locations for the PICRT stations and the allocation of DGHs together, the optimisation is split into two stages, where we assume the number and location of PICRT stations are given and solve for them the optimal allocation of DGHs, and afterwards optimise the configuration of PICRT stations. This allows us to deconstruct a large problem into several manageable parts. Both problems will be solved by applying a genetic algorithm.

3.5. Part 1: given PICRT station locations and their number

Given a set of operational PICRT stations, we can reduce the number of variables Y_{ji}, since there are now only N number of possible values of j. The objective function in equation (4) remains the same but only the terms where station j is an operational PICRT station remain. The constraints (C1), (C2), and (C4) will be imposed, whereas (C3) does not matter at this stage because the operational PICRT stations are already known.

There is another constraint that can be imposed on the variables based on the idea that the practical solution must be one for which each station is able to satisfy the demand of the DGHs allocated to it and not be overloaded, that is rate of new requests should not exceed the rate at which they can be served. Mathematically, the requirement is expressed as

$$(\text{C5})\ \rho_j = \frac{\lambda_j}{c_j\mu_j} = \sum{}_{\cdot i}d_iY_{ji}\mathbb{E}[T_{ji}] \le 1. \tag{5}$$

In summary, when the set of operational PICRT stations are given, we solve the problem

$$\min_{Y_{ji}} \sum{}_{\cdot j}W_j(\{Y_{ji}\}) + \sum{}_{\cdot i,j}d_iY_{ij}(w_1(T_m + t_{ij}^1 + T_a) + w_2(T_p + t_{r(i)i}^2) + w_3t_{jr(i)}^3)$$

$$\text{subject to (C1), (C2), and (C5).} \tag{6}$$

It is possible that this has no feasible points, in which case we arbitrarily assign the answer to be an extremely high number, say 10^{10}.

3.6. Part 2: Optimising PICRT station locations

We now have a map from the variables X_j's to the minimum value obtained by solving equation 6 (the allocation of DGHs to PICRT locations j), which we write as $F(X_j)$. The resulting minimisation problem is

$$\min_{\{X_j\}} F(\{X_j\}) \text{subject to (C3).} \tag{7}$$

These two minimisation problems working together yield the optimal location of PICRT stations and allocation of DGHs.

3.7. Distribution of teams among selected locations

The final important problem is to assign the optimal number of teams to each operational station given an overall number of teams. A natural solution is to consider c_j as a decision variable, instead of a parameter, along with the optimal operating stations and hospital allocations and apply the genetic algorithm.

However, due to the nonlinearity of the objective function and the size of our problem, we used a simpler method to arrive at team distributions to maintain practical computational solution times. Our strategy is to over-prescribe teams to each station and use this over-prescribed team profile to solve the objective function given in equation 4. Then, using the obtained retrieval team locations and hospital allocations, teams are sequentially removed from each station using a greedy algorithm; that is, a team is removed if its removal leads to the least increase in the optimised objective function value.

Let C denote the number of teams to be distributed and let c' be the number of teams prescribed to each station. c' is chosen such that $c'N >= C$, where N is the number of operational stations. After obtaining the operational station locations and the hospital allocations, the greedy algorithm is employed. The choice of c', however, can affect the result of its corresponding stations and hospital allocations obtained from equation (4). For example, the higher value of c', the closer the solution would be to the linear optimal solution (without wait time). Therefore, we perform the optimisation for several values of c', apply the greedy algorithm, and choose the resulting solution that yields the lowest objective value.

3.8. Software

The model was coded in MATLAB and solved using *intlinprog* for integer linear programs and *ga* for the integer genetic algorithm. The Google Maps Distance Matrix API is used to obtain the mean travel times not obtainable from historical PICRT transport data. The Google Maps Geocoding API is used to plot the allocations through the *gmplot* module in Python.

3.9. Ethical approval

DEPICT has ethical approval from the Health Research Authority, the National Research Ethics Service (London Riverside, reference: 17/LO/1267) and the Confidentiality Advisory Group (reference: 17CAG0129).

4. Results

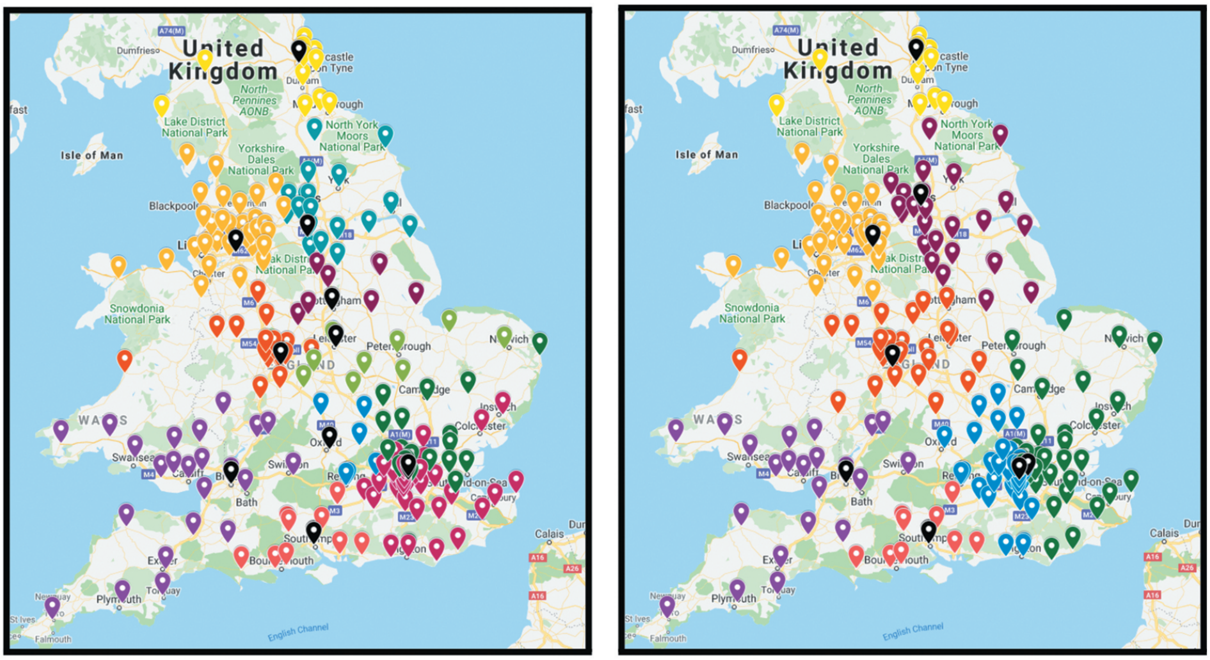
We applied our optimisation framework to England and Wales to obtain configurations of operational PICRT stations as well as an allocation of local hospitals to these stations across a range of scenarios. We describe how average time to bedside, waiting time, and the percentage of demand covered differ with demand load, number of teams in service, and number of operational PICRT stations. The parameterisation and scenarios considered were chosen together with our clinical collaborators. We received pseudonymised data from the Paediatric Intensive Care Audit Network (PICANet) on all journeys by PICRT services between 2014 and 2018. This comprised over 15,000 transports which were used to estimate: demand for service from each DGH, mean and variance of journey times between hospitals, and the proportion of each hospital's demand that went to each receiving PICU. Where no journeys between a pair of hospitals was recorded (e.g., for instance, a hospital in the South would not be served by a PICRT in the North and so we would not expect a journey between the two) we estimated the mean travel times using Google Maps.

In the current service, there are 212 hospitals, 24 PICUs, and 11 operating PICRT stations. We allow any PICU to be a potential location for a PICRT station, giving us 28 possible PICRT locations (24 PICUs plus 4 further locations which currently host PICRTs). The 212 DGHs are hospitals, which may include some PICUs, that have used the PICRT service at least once from 2014 to 2018.

The data were fed into a genetic algorithm as described in the previous section. The search for the optimal allocation was done with an initial population of 50 configurations and was run over 20 generations. The search for the optimal PICRT station configuration had an initial population of 50 configurations and was run over 10 generations.

4.1. Changing number of PICRT locations

We illustrate the optimisation framework given in equation (4) by comparing results for 8 PICRT stations with 11 PICRT stations (see Figure 2), with 16 PICRTs in total and travel leg weights of $w_1 = w_2 = 2$ and $w_3 = 1$. This is to reflect that fact that getting a child to the receiving PICU is considered more important than returning to base quickly once the child has been transferred. When the number of stations is reduced from eleven to eight, one station is removed from London, the remaining two London stations dividing the service on an East-West axis. Two further stations are removed from the Midlands and Cambridge. Other allocations remain roughly the same because hospitals in those regions such as the Manchester and Newcastle cluster are separated far



(a) 11 stations 16 teams, $w_1 = w_2 = 2, w_3 = 1$ (b) 8 stations 16 teams, $w_1 = w_2 = 2, w_3 = 1$

Figure 2. DGH Allocation under different number of stations. With fewer stations, the Midlands, the East and London are served by one fewer station each.

enough from other PICRT stations that any changes to their current situation will incur too great a travel time cost.

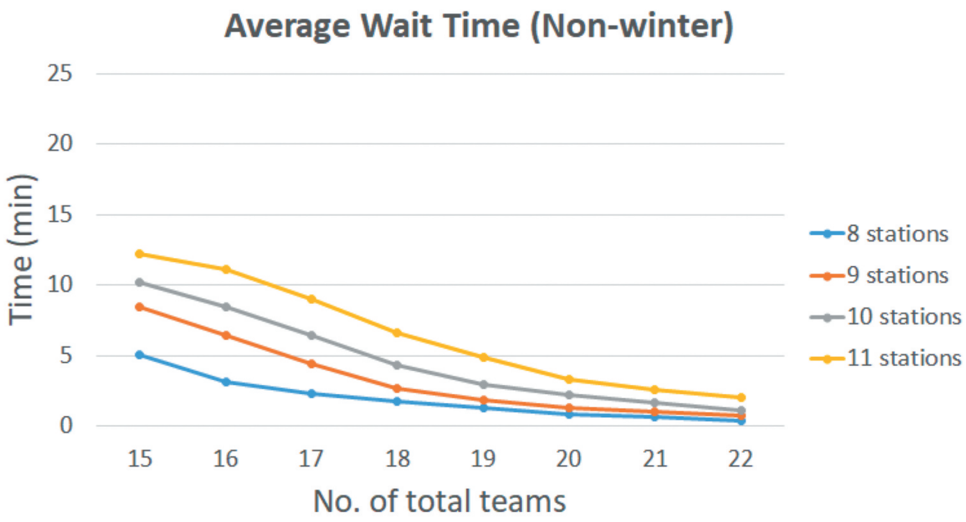
Using the average historical demand for the retrieval service during the winter, the calculated expected time to bedside and time to PICU are 97.9 minutes and 281.6 minutes, respectively, when using 8 stations and 110.4 minutes and 294.1 minutes, respectively, when using 16 stations. The advantage of using eight stations here is that this configuration trades geographical proximity with the availability of a retrieval team. When 16 teams are spread out, the waiting period for a next available team increases, particularly at times of high demand. Allowing more than one team per station reduces this wait time at a relatively lower cost of travel time.

4.2. Comparing solutions with changing demand and number of teams

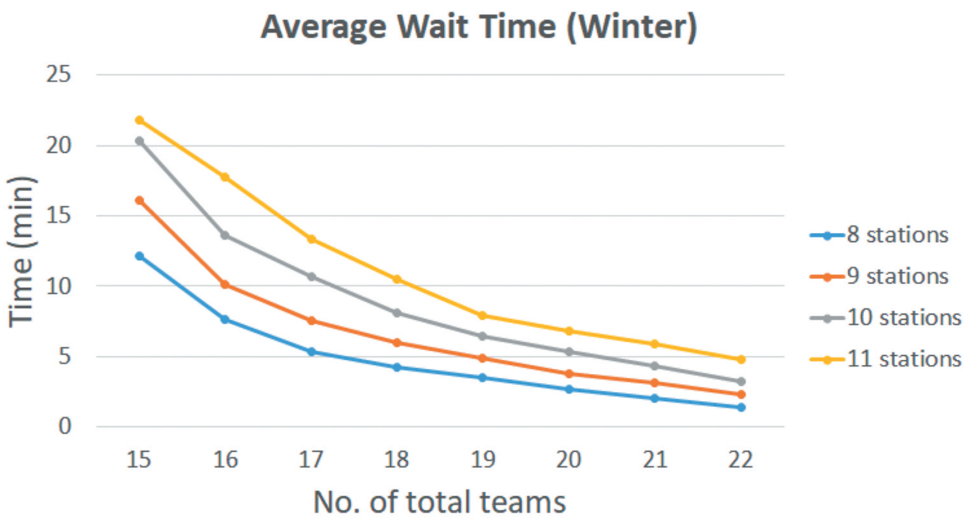
To compare solutions obtained from our extended model more broadly, we consider two metrics: the average waiting time (the time for the next PICRT to be available) and the average time to bedside. After discussion with clinical partners, each metric is evaluated with 15 to 22 teams for between 8 and 11 stations (there are currently typically 16 teams across the country during daytime). We compare the performance between two levels of demand calculated from PICANet data: one is the demand during the day (8am-8pm) in the months March to October, which we will call non-winter, and one during the day in the months November to February, which is winter (when demand typically doubles). We concentrated on day time since we wanted to optimise the service for the busier part of the 24-hour day.

In Figure 3, we show how average waiting time for an available PICRT changes with total number of teams and number of locations (stations) for winter (Figure 3b) and non-winter (Figure 3a). Because the demand for the PICRT service is higher during the winter, the waiting time for a team must correspondingly be greater. Waiting time reduces as more teams are available. It is generally more efficient to have the same number of teams spread across fewer locations as it is then more likely that a team is available at a given location (e.g., waiting time for 16 teams across 11 stations is more than twice as high as for the same number of teams across 8 stations).

In Figure 4 we show the corresponding times to bedside for winter (Figure 4b) and non-winter (Figure 4a). Again, the time to bedside is greater for the demand during winter and generally decreases as more teams are used. However, there is an interesting trade-off between the efficiency in waiting time for teams across fewer locations and the fact that fewer locations means that, inevitably, some DGHs will be further away with correspondingly longer travel times. This trade-off is more pronounced in winter. For a lower number of teams, fewer stations is better as teams are a scarce resource and reducing waiting time becomes more important than reducing journey time. However, after 18 or 19 teams, there are sufficient teams to provide a short waiting time, and journey times become more important. So while 8 is the optimal number of locations for 15 teams it is the worst number of locations



(a) Avg. Waiting Time (day, summer), $w_1 = w_2 = 2, w_3 = 1$



(b) Avg. Waiting Time (day, winter), $w_1 = w_2 = 2, w_3 = 1$

Figure 3. Average waiting time with respect to 8 to 11 PICRT stations and 15 to 22 total PICRTs with (a) average demand on a summer's day and (b) average demand on a winter's day. The waiting time is greater for winter, which has a higher demand for PICRT service, decreases with number of teams and increases with number of stations.

once you have 19 teams available. 9 or 10 stations are instead optimal, with little improvement in journey times after 19 teams.

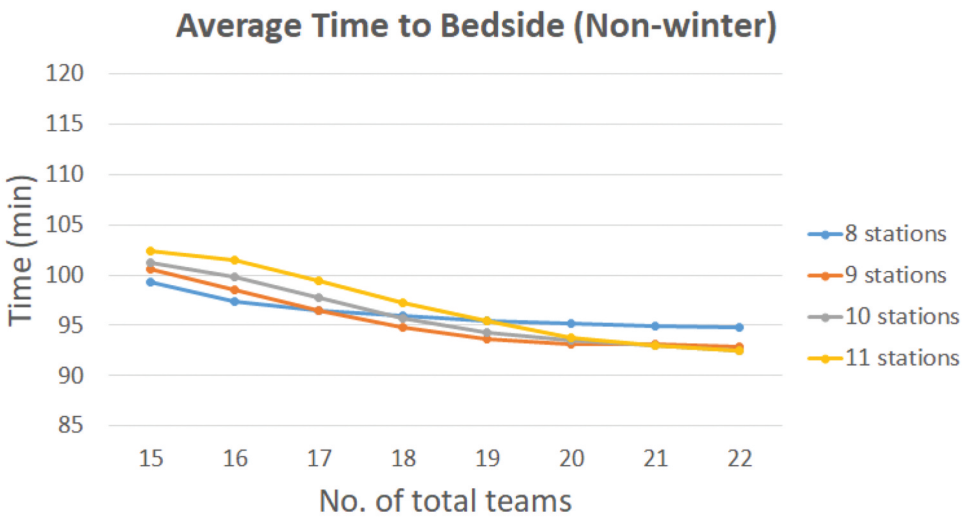
We also see from Figure 4 that for the same number of teams available, time to bedside is higher in winter than out of winter (unsurprisingly, given increases in waiting time under higher demand). However, adding teams in winter could provide similar times to bedside as in non-winter: for instance, if 15 teams were available across 9 locations in non-winter, extra 4 teams in winter across 9 locations would achieve similar time to bedside.

5. Limitations

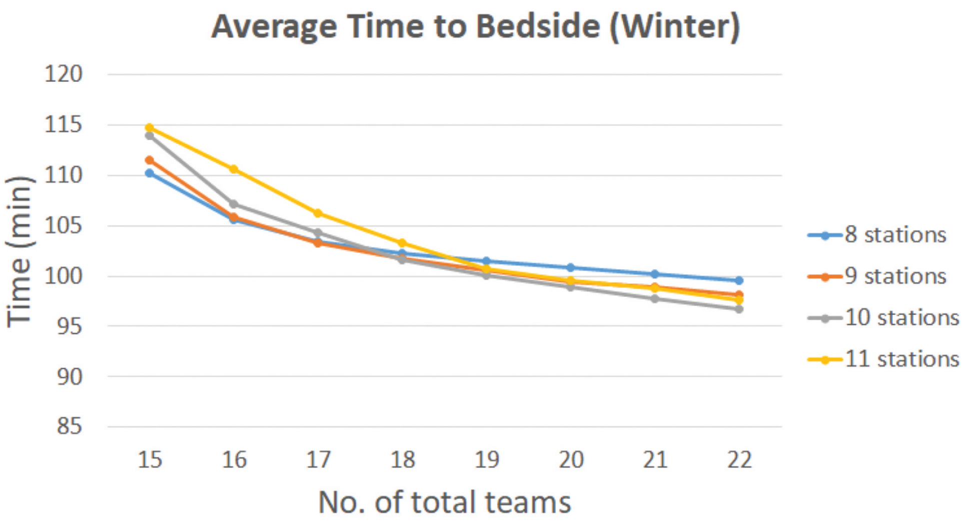
While our model has generated several useful insights, there are some limitations.

We currently assume that the mobilisation time, time for the team to get to the bedside from ambulance arrival at the DGH, and the treatment time at the bedside before transport to the PICU to be constant and the same for all locations. The mobilisation time and time to reach the child's bedside are small compared to journey time but the time at bedside can be long (several hours) and quite variable, depending on the condition of the child. However, for informing decisions about location of services and number of teams at each service, which will depend most on demand and journey times, treating these variables as constant (using observed mean values) is reasonable.

The Kingman's formula approximation presents the main limitation to our model. One problem posed by the approximation is the assumption of steady state. Depending on how fast a given scenario



(a) Avg. Time to Bedside (day, summer), $w_1 = w_2 = 2, w_3 = 1$



(b) Avg. Time to Bedside (day, winter), $w_1 = w_2 = 2, w_3 = 1$

Figure 4. Average time to bedside with respect to 8 to 11 PICRT stations and 15 to 22 total PICRTs with (a) average demand on a summer's day and (b) average demand on a winter's day. The time to bedside decreases with the number of teams. Fewer stations performs comparatively better when fewer teams are available while more stations performs comparatively better with more teams.

converges to the steady state and the volatility of service demand, it may be reasonable to make this assumption and one could argue that assuming steady state is sensible for decisions on long-term service provision. However, if the system does not converge quickly, this assumption may not hold. In that case, the waiting time could be obtained using alternative methods or the framework must be adapted to the change in demand throughout the year. Furthermore, the heuristic approach to solve the non-linearity introduced by the formula cannot guarantee the solution to be globally optimal.

Another limitation is the assumption of a First Come First Served protocol for attending new referrals. While this is normally the case, where two referrals come in at once and only one team is available, the sicker child is prioritised. We have not incorporated such complexity into our framework.

Finally, we have set up a simplified way to determine the number of teams to be allocated to each PICRT location. Having the distribution of teams included as a variable to the optimisation greatly increases the size of the problem and the computational times to solve it are not practical. However, future work will include designing a more efficient algorithm that incorporates the distribution of teams as a variable along with the other variables.

6. Discussion

We have developed an optimisation framework that incorporates stochastic journey times, team availability and different levels of demand for paediatric intensive care retrieval services from each local district general hospital (DGH). We have used the framework to explore different scenarios with clinical collaborators for different numbers of locations, different overall numbers of teams available, and different levels of seasonal and daytime demand.

Exploring the various scenarios showed that number of locations is less important than total number of teams, particularly when more teams were available. For a smaller number of teams, fewer locations tends to be better as more locations can then "double up" on teams. This is consistent with the general principle from queuing theory that it is better to have more servers and a single queue (more teams at one location) than many queues with a single server (more locations with one team each). Another important result is that for a given number of locations, we can inform decisions to increase team capacity in the busy winter surge to keep times to bedside the same as for non-winter on average. Not only can we say how many more teams would be needed, but we can also inform where they should be allocated. Unsurprisingly, the locations with the highest demand tend to be those that are allocated the extra teams.

The other insight from the piece is that many different configurations of numbers of locations and teams give very similar performance in terms of time to bedside – within about 10 min of each other. Our next step is to work with our clinical collaborators to identify configurations that provide both good performance in terms of time to bedside, during both winter and non-winter, and are also pragmatic to implement given existing infrastructure. For instance, it would not be sensible to advocate changing many locations (and the fixed costs associated with such a move) for a marginal gain of a few minutes. This process will also be informed by the findings from the other strands of the DEPICT project such as family experience (Evans et al., 2019). Finally, the successful development of this framework in close collaboration with clinical teams, and with interest from policy makers in the national Paediatric Intensive Care Society, should lead to increased use of these operational research methods to inform not just future configuration options for PICRT services but also future decisions about the further centralisation (or not) of paediatric intensive care service more generally.

Disclosure statement

No potential conflict of interest was reported by the authors.

Funding

This work was supported by the UK National Institute for Health Research, Health Services and Delivery Research programme (NIHR, HSDR) [Project No:15/136/45].

References

Allen, A. O. (1990). Probability, Statistics, and Queueing Theory: With Computer Science Applications. Academic Press Professional, San Diego, USA. ISBN-13 978-0120510511

Shiah DM & Chen, SW. Ambulance allocation capacity model. IEEE 9th International Conference on e-Health Networking, Application and Services, Taipei, Taiwan, https://ieeexplore.ieee.org/document/4265794, 2007.

Church, R., & ReVelle, C. (1989). The maximum availability location problem. *Transportation Science*, *23*(3), 192–200. https://doi.org/10.1287/trsc.23.3.192

Coates, G., & McCormack, R. (2015). A simulation model to enable the optimization of ambulance fleet allocation and base station location for increased patient survival. *European Journal of Operational Research*, *247*(1), 294–309. https://doi.org/10.1016/j.ejor.2015.05.040

Dale, D. (1979). Achabal and Milton E.F. Schoeman. An examination of alternative emergency ambulance systems: Contributions from an economic geography perspective. *Social Science & Medicine. Part D, Medical Geography*, *13*(2), 81–86. doi:10.1016/0160-8002(79)90054-6

Daskin, M. S. (2008). What you should know about location modeling. *Naval Research Logistics (NRL)*, *55*(4), 283–294. https://doi.org/10.1002/nav.20284

Evans, R., Draper, E. S., Seaton, S. E., Wray, J., P., S. M. R., & Pagel, C. (2019). Differences in access to emergency paediatric intensive care and care during transport (depict): Study protocol for a mixed methods study. *BMJ Open*, 9: e028000. https://doi.org/10.1136/bmjopen–2018–028000

Ingolfsson, A., Sim, T., Erdogan, G., & Erkut, E. (2009). Computational comparison of five maximal covering models for locating ambulances. *Geographical Analysis*, *41*(1), 43–65. https://doi.org/10.1111/j.1538-4632.2009.00747.x

King M, Ramnarayan, P., Seaton, S.,Pagel, C. (2019). Modelling the allocation of paediatric intensive care retrieval teams in england and wales. *Archives of Disease in Childhood*, *104*(10), 962. https://doi.org/10.1136/archdischild-2018-316056

Kingman, J. F. C. (1961). The single server queue in heavy traffic. *Mathematical Proceedings of the Cambridge Philosophical Society*, *57*(4), 902. https://doi.org/10.1017/S0305004100036094

Larson, R. C. (1974). A hypercube queuing model for facility location and redistricting in urban emergency services. *Computers & Operations Research*, *1*(1), 67–95. https://doi.org/10.1016/0305-0548(74)90076-8

Li, X., Zhao, Z., Zhu, X., Wyatt, T. (2011). Covering models and optimization techniques for emergency response facility location and planning: A review. *Mathematical Methods of Operations Research*, *74*(3), 281–310. https://doi.org/10.1007/s00186-011-0363-4

Burkey, M., Bhadury, J., & Eiselt, H. A. (2012). A location-based comparison of health care services in four u.s. states with efficiency and equity. *Socio-Economic Planning Sciences, Modeling Public Sector Facility Location Problems*, *46*(2), 157–163. doi: 10.1016/j.seps.2012.01.002

Mirchandani, P. B., & Odoni, A. R. (1979). A simulation model to enable the optimization of ambulance fleet allocation and base station location for increased patient survival. *Transportation Science*, *13*(2), 85–97. https://doi.org/10.1287/trsc.13.2.85

Mitchell, M. (1990). *An Introduction to Genetic Algorithms*. MIT Press.

Knight, V. A., Harper P.R., & Smith, L. (2012). Ambulance allocation for maximal survival with heterogeneous outcome measures. *Omega, Special Issue on Forecasting in Management Science*, *40*(6), 918–926. doi: 10.1016/j.omega.2012.02.003

Paediatric Intensive Care Audit Network. *Paediatric Intensive Care Audit Network Annual Report (2018) Universities of Leeds and Leicester*. http://www.picanet.org.uk/, 2020.

Paediatric Intensive Care Society. (2015) . *PICS. Quality Standards for the Care of Critically Ill Children, 5th edition*.

Parslow, R. C., Harrison, D. A., P., E. S. D. R., Rowan, K. M., & Rowan, K. M. (2010). Effect of specialist retrieval teams on outcomes in children admitted to paediatric intensive care units in england and wales: A retrospective cohort study. *Lancet*, *376*(9742), 698–704. https://doi.org/10.1016/S0140-6736(10)61113-0

Berman, O., Larson, C., Parkan, C. (1987). The Stochastic Queue p-Median Problem. *Transportation Science*, *21*(3), 207–216. https://doi.org/10.1287/trsc.21.3.207

Ramnarayan S., P., Ray Pagel, C., Peters, M. J., & Peters, M. J. (2015). Novel method to identify the start and end of the winter surge in demand for pediatric intensive care in real time. *Pediatric Critical Care Medicine*, *16*(9), 821–827. https://doi.org/10.1097/PCC.0000000000000540

Ramnarayan S., P., Ray Pagel, C., Peters, M. J., & Peters, M. J. (2018). Development and implementation of a real time statistical control method to identify the start and end of the winter surge in demand for paediatric intensive care. *European Journal of Operational Research*, *264*(3), 847–858. https://doi.org/10.1016/j.ejor.2016.08.023

Silva, F., & Serra, D. (2016). Locating emergency services with different priorities: The priority queuing covering location problem. *Operational Research for Emergency Planning in Healthcare: Volume 1, Edited by Navonil Mustafee*, 15–35. doi: 10.1057/9781137535696_2

Toregas, C., Swain, R., ReVelle, C., & Bergman, L. (1971). The location of emergency service facilities. *Operations Research*, *19*(6), 1363–1373. https://doi.org/10.1287/opre.19.6.1363