Highlights

Evaluating Heterogeneous Ambulance Fleet Allocations in Jakarta

Geraint Palmer, Mark Tuson, Sarie Brice, Paul Harper, Vincent Knight, Leanne Smith, Daniel Gartner

- Mathematical modelling for maximising survival with heterogeneous patients and fleets
- Heuristic algorithm for finding better fleet allocations
- A sequential simulation for ambulance demand and capacity planning
- Application to Jakarta for policy recommendations and testing future demand scenarios

Evaluating Heterogeneous Ambulance Fleet Allocations in Jakarta

Geraint Palmer[a], Mark Tuson[a], Sarie Brice[a], Paul Harper[a], Vincent Knight[a], Leanne Smith[b], Daniel Gartner[a,c]

[a]School of Mathematics, Cardiff University, Abacws, Senghennydd Road, Cardiff, CF24 4AG, Wales, United Kingdom
[b]Welsh Ambulance Services NHS Trust, Beacon House, William Brown Close, Cwmbran, NP44 3AB, Wales, United Kingdom
[c]Aneurin Bevan University Health Board, Lodge Road, Caerleon, NP18 3XQ, Wales, United Kingdom

Abstract

Ambulance services have a duty of care to the clinical outcomes of the population they serve, and therefore aim to maximise the chances of survival and improve patient outcomes following a medical emergency. Despite this, although the ambulance location-allocation problem has been widely studied, it has predominantly focused on minimising response times or maximising coverage alone, and not explicitly for considering patient outcomes. In this paper we propose a modelling approach to consider where to optimally locate different types of emergency response vehicles in order to maximise patient outcomes within a heterogeneous population. To achieve this, we develop a heuristic algorithm for finding better fleet allocations which is used in conjunction with a discrete-event simulation model of ambulance services with heterogeneous vehicles. Our approach is informed by, and tested on, real-world data from Jakarta, Indonesia. Using our developed models, decision makers are better able to understand ambulance fleet capacity needs and locations, and their impact on patient outcomes.

Keywords: Ambulance Planning, Emergency Medical Services, Simulation, Health Care, Optimisation, Mathematical Modelling

1. Introduction and Background

Time-critical conditions (TCCs) are a substantial cause of mortality worldwide, responsible for an estimated 54% of all deaths [13]. Emergency medical services (EMS) play a pivotal role in responding to TCCs and saving patients from life-threatening medical conditions. Most EMS research, especially planning to meet increasing ambulance demands, tends to be focussed on high-income countries including, for example, Australia [22], Germany [35], and USA [5]. In contrast, most low and middle-income countries (LMICs) lack an organised EMS system, with most ambulances used purely for patient transport and not as an emergency care vehicle [28]. However, the burden of deaths due to TCCs is much greater in LMICs than in high-income countries, the difference being around threefold [9]. For traffic accidents, strokes and heart attacks, rapid access to appropriate treatment is especially vital.

The application of our research is focused on Indonesia, as supported by an award for Global Challenges Research Funds (GCRF). In Indonesia, an LMIC, there has been very little prior work on developing an EMS strategy, driven largely by inadequate resources including a lack of financial investment and human capital [28, 29, 37]. Related research by the authors [8] has highlighted particular challenges in Indonesia as a lack of a single coordinated EMS, the ability or willingness to pay for an ambulance, a large geographical area, and areas of severe traffic congestion especially in Jakarta, the capital.

When our research programme commenced in October 2019, ambulance services in Jakarta were provided by many disparate, mostly private providers that charge patients for their use, such as private hospitals. We initially partnered with Ambulans 118, a non-government charitable ambulance service established in 2005 by the Indonesian Surgeons' Association, that currently operates in five cities across Indonesia: Jakarta, Palembang, Yogyakarta, Surabaya and Makasser. Unlike private providers, Ambulans 118 suggests that a donation is made for use of its EMS vehicles, but otherwise provides free emergency medical care and is also leading on paramedic training across the country.

The overarching goal of our research was to work with the Indonesian Government, the different ambulance providers, and hospitals, to help them forecast emergency demand and make critical decisions on the optimal types, capacities and geographical locations of emergency vehicles within a potentially co-ordinated EMS system, starting with Jakarta. To facilitate this, in collaboration with Ambulans 118 we organised workshops in Jakarta in

February 2020, thankfully just before the COVID-19 pandemic and lockdowns, which were attended by over 170 people including doctors, nurses, paramedics, academics, and officials from the Indonesian Ministry of Health. This paper presents some aspects of the overall research programme, specifically focusing on the development of a simulation model and location-allocation heuristics for maximising patient survival.

Fundamental to the project and the research described in this paper was an initial study that we carried out in order to better understand patient needs and the barriers to use of ambulances in Indonesia. Throughout the month of December 2020, we undertook comprehensive surveys in Emergency Departments (EDs) across Jakarta and published the first known study in to EMS demand within the country [8]. Our study showed that the utilisation of ambulances by patients attending EDs is considerably low and of concern. The low utilisation is contributed to patients' lack of awareness of available ambulance services given a lack of coordination, patients' disinclination to use ambulances due to high costs, and long response times. All of these barriers impact on patient outcomes, especially for those with life-threatening conditions. For example, remarkably more trauma patients took a car-share ride (20%) or motorcycle (20%) to reach the ED than an ambulance (just 10%), while only 14% of critical cardiovascular patients used an ambulance compared to 67% travelling to hospital by private car or a car-share ride.

The first contribution of this paper is the development of a comprehensive sequential discrete event EMS simulation that models and evaluates heterogeneous fleet allocations of emergency vehicles by utilising a novel approach in which transit jobs, rather than patients, are framed as the queueing 'customers'. Our methodology utilises two discrete event simulations of the same system that are run sequentially and together combine to form the logic of a single simulation of a heterogeneous fleet; and key performance indicators (KPIs) are calculated by making use of survival functions. Secondly, we extend the work of [19] to consider both heterogeneous patients and heterogeneous fleets, utilising a cross-entropy heuristic to find allocations that maximise patient survival. The resulting allocations are evaluated using the discrete event simulation. Thirdly, we demonstrate the use and impact of our modelling framework applied to current and proposed ambulance allocations in the city of Jakarta, thus supporting Government-level decision making with an overall goal to improve the lives of those living in LMICs. This leads to decision support and managerial insights in Indonesia, although the developed modelling framework could be readily applied to other locations.

The paper is structured as follows: Section 2 gives an overview of the literature on modelling ambulance allocations. Section 5 outlines the logic of the sequential discrete event simulation models for heterogeneous fleets. Section 6 describes the current emergency service situation in Jakarta and presents simulation results from current and proposed fleet allocations. Section 4 describes the survival objective and heuristic algorithm to find better fleet allocations, and Section 7 discusses the findings and contributions of the paper.

2. Literature Review

There is a rich literature on Operational Research applied to EMS location and related allocation problems. Literature reviews that include these topics are Aringhieri et al. [2], Bélanger et al. [4], Farahani et al. [12], Li et al. [20], Liu et al. [21], Reuter-Oppermann et al. [30] and Wang et al. [36].

Emergency medical services can be evaluated using different performance measures with coverage and response time being the predominant metrics [24]. However, missing a response time target by just one second, for example, would be considered as a 'failure' in many models with no appreciation at all of the impact on patient survival or outcome [24]. This seems somewhat restrictive and short-sighted, which is why we focus on survival as a metric and thus, in the following, compare and contrast papers using that metric with our approach.

Table 1 provides an overview of recent research that focus on the survival metric. We break down related papers into discrete-event simulation (DES), queueing theory and optimisation methods, and we provide a summary of each paper from the table as well as compare and contrast them with our modelling and solution approach.

Reference	Year	DES	Queueing Theory	Optimisation
Amorim et al. [1]	2019			✓
Boutilier and Chan [6]	2022		✓	
Erkut et al. [11]	2008			✓
Knight et al. [19]	2012		✓	✓
McCormack and Coates [23]	2015	✓		✓

Table 1: Break down of related research that focuses on survival.

Amorim et al. [1] propose an integrated strategic and tactical planning approach which features an optimization model and a local search heuristic based on Gaussian Processes. Their methodology is applied to the city of Porto while reporting on performance metrics such as survival. Our approach is different because we use a discrete-event simulation framework and a cross-entropy optimisation approach to determine the number of ambulances required in each station.

Boutilier and Chan [6] develop an integrated location-queueing model that incorporates existing EMS response times in a drone network. They use a p-median approach and an Erlang loss model. Although survival is not their main optimization criterion, they report how many patients would survive using their approach.

Erkut et al. [11] can be considered the first modelling approach that considers decaying survival probabilities during response time. The rationale is, because survival can be thought of as a more robust and generic objective for EMS performance measurement than coverage or average response time. In an application of a recently developed Maximal Survival Location Problem model (MSLP), the authors use data from Edmonton, Canada and show that maximising survival is superior to other objectives in clinical outcome for cardiac arrest patients.

McCormack and Coates (2015) [23] focus on the optimization of EMS vehicle fleet allocation and base station location through the use of a genetic algorithm (GA) with an integrated EMS simulation model. Their objective is maximization of the overall expected survival probability across patient classes. Applications of the model were undertaken using real call data from the London Ambulance Service. The difference between their modelling approach and our is that we have a different survival function, focus on a different heuristic optimization and evaluate our approach with data from a development country and have higher traffic volume and congestion in our data.

Knight, Harper and Smith [19] provide a supporting extension of the MSLP that allows it to be applied more generally to real-world EMS systems, acknowledging that different patients have varying levels of expected survival probabilities. The Maximum Expected Survival Location Model for Heterogeneous Patients, MESLMHP, allows multiple classes of patient groups to be defined, where previously only cardiac arrest patients formed part of the objectives for primary response. In reality, by definition any emergency patient has a necessity for swift attendance, and a timely response to

any incident type may impact on clinical outcome in some way. It is for this reason MESLMHP is designed to be generic enough to accommodate any number of patient groups, with each class weighted dependent upon the relative urgency of the incident. More than just a contribution to the model's demand input, these patients are included in the optimisation when maximising total population survival probability.

3. Survival Functions

A key concept here is the survival curve of a patient. In reality, some emergency incidents do not result in a substantive deterioration of a patient's status over time; however, in all situations, there is a reasonable cut-off beyond which a patient should not expect to have to wait for care, and mixture of theoretical survival functions and step functions can be utilised. Survival probabilities for critical incidents, calculated from a theoretical monotonically decaying survival function reported in the literature [34] are used to demonstrate an attainable level of success from a response. One particular survival curve $s_1(t)$ of Equation 1, represents survival until hospital discharge following cardiac arrest; its origins are explained in detail by [19]. It is used here for critical A1 patients. Figure 1 shows the difference between using this survival curve and a hard cut-off of 8 minutes. However, hard cut-off curves like Equation 2 can still be used to represent meeting artificially selected targets and are used here for the survival of A2 and B patients, with targets of 15 and 60 minutes respectively, given by Equations 3 and 4.

$$s_1(t) = \left(1 + e^{0.26+0.139t}\right)^{-1} \tag{1}$$

$$s_2(t) = \begin{cases} 1 \text{ if } 0 \leq t \leq 8 \\ 0 \text{ if } t > 8 \end{cases} \tag{2}$$

$$s_3(t) = \begin{cases} 1 \text{ if } 0 \leq t \leq 15 \\ 0 \text{ if } t > 15 \end{cases} \tag{3}$$

$$s_4(t) = \begin{cases} 1 \text{ if } 0 \leq t \leq 60 \\ 0 \text{ if } t > 60 \end{cases} \tag{4}$$

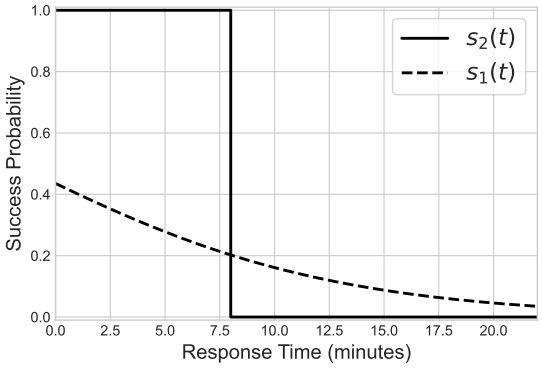


Figure 1: Survival function $s_1(t)$, estimated by [34] compared with a current EMS target, $s_2(t)$, represented as a step function (binary coverage).

4. Finding Better Allocations

In this section we propose a heuristic approach of finding ambulance allocations for a given resource level. A cross-entropy optimisation approach is used to find allocations that maximise the expected survival of the patients served. These allocations are also simulated using the developed simulation described in Section 5 and demonstrated in Section 6, to obtain their equivalent simulation based KPIs.

First, in Section 4.1 a maximal expected survival score is given for heterogeneous patients and heterogeneous vehicles, to be used an objective function. Then in Section 4.2 a cross-entropy heuristic algorithm is described to find better fleet allocations, while Section 6.5 describes the design of the numerical experiments across the relevant scenarios. Section 6.6 provides the results of the optimisation based on the KPIs obtained through the discrete event simulation.

4.1. MESLMHPHF Objective Function

The intent of the objective function used for the heuristic optimisation algorithm is to maximise patient survival. Here we formulate a weighted expected survival function as the objective function which considers heterogeneous patients (different specialities) and heterogeneous fleets (both primary vehicles, EAs, and secondary vehicles, RRVs), called the Maximal Expected Survival Location Model with Heterogeneous Patients and Heterogeneous

Fleet (MESLMHPHF). It is an extension of the MESLMHP model given in [19], which did not consider heterogeneous fleets.

Our maximal survival function is constructed by appropriately summing these survival curves, described in Section 3, across the population, multiplying by ambulance availability where appropriate. We let Z_a and $\tilde{Z}_a$ denote the number of primary vehicles and secondary vehicles at ambulance station a, respectively, that is, the allocation. To incorporate heterogeneous fleets, as was modelled by the simulation in Section 5, the set of specialities $\mathscr{K}$ is partitioned into two sets, $\mathscr{K}_A$, those patients that can be seen by secondary vehicles (in the Jakarta case including specialities A1 and A2), and $\mathscr{K}_B$, those patients who are seen by primary vehicles only (in the Jakarta case including speciality B only). Then, the MESLMHPHF objective function is given by Equation 5,

$$g\left(Z_a, \tilde{Z}_a\right) = \sum_{p\in\mathscr{P}} \sum_{a\in\mathscr{A}} \left(\sum_{k\in\mathscr{K}_A} w_k \lambda_{pk} \hat{\Psi}_{kpa} + \sum_{k\in\mathscr{K}_B} w_k \lambda_{pk} \Psi_{kpa} \right) \tag{5}$$

where w_k is a weight associated with patient speciality type k (for this study we assume $w_k = 1$ for all k), λ_{pk} is the rate of demand of patients of speciality k at pick-up location p. Now Ψ_{kpa} is the expected survival of patients of speciality $k \in \mathscr{K}_B$ at location p from station a, while $\hat{\Psi}_{kpa}$ is the expected survival of patients of speciality $k \in \mathscr{K}_A$ at location p from station a. Equation 5 is the weighted sum over the expected survival functions of all patient specialities, all patient pick-up locations, and all ambulance stations. These expected survival functions are given by Equations 6 and 7 respectively,

$$\Psi_{kpa} = s_k\left(t_{pa}\right)\left(1-\pi_a^{Z_a}\right) \prod_{\alpha\in\mathscr{A}} \pi_\alpha^{\left(Z_\alpha \beta_{p\alpha a}\right)} \tag{6}$$

$$\begin{aligned} \hat{\Psi}_{kpa} = {} & s_k\left(\hat{t}_{pa}\right)\left(1-\tilde{\pi}_a^{\tilde{Z}_a}\right) \prod_{\alpha\in\mathscr{A}} \tilde{\pi}_\alpha^{\left(\tilde{Z}_\alpha \beta_{p\alpha a}\right)} \pi_\alpha^{\left(Z_\alpha R_{p\alpha a}\right)} \\ & + s_k\left(t_{pa}\right)\left(1-\pi_a^{Z_a}\right) \prod_{\alpha\in\mathscr{A}} \pi_\alpha^{\left(Z_\alpha \beta_{p\alpha a}\right)} \tilde{\pi}_\alpha^{\left(\tilde{Z}_\alpha\left(1-R_{p\alpha a}\right)\right)} \end{aligned} \tag{7}$$

Here s_k denotes the survival functions of patients of speciality k (Equation 1 when k represents A1 patients, and Equations 3 and 4 when k represents A2 and B patients respectively); t_{pa} and $\hat{t}_{pa}$ denote the expected time

to travel from a to p for primary vehicles and secondary vehicles respectively; and π_a and $\tilde{\pi}_a$ is the utilisation at ambulance station a of primary and secondary vehicles respectively. $R_{pa_1a_2}$ and $\beta_{pa_1a_2}$ are binary variables indicating the preference of sending a vehicle from one station to another: $\beta_{pa_1a_2}$ indicates if a vehicle of the same type can reach p quicker from a_1 than a_2, defined in equation 8, while $R_{pa_1a_2}$ indicates if a primary vehicle at a_1 can reach p quicker than a secondary vehicle at a_2, defined in equation 9.

$$\beta_{pa_1a_2} = \begin{cases} 0 & \text{if } a_1 = a_2 \\ 1 & \text{if } t_{pa_1} \leq t_{pa_2} \\ 0 & \text{otherwise.} \end{cases} \tag{8}$$

$$R_{pa_1a_2} = \begin{cases} 1 & \text{if } t_{pa_1} \leq \hat{t}_{pa_2} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

Equation 6 is the probability of a patient surviving (their survival function), multiplied by the availability of primary vehicles at that ambulance station, multiplied by the unavailability of vehicles at closer stations. Equation 7 extends this to two vehicles types, the first part of the sum repeating the logic for the faster secondary vehicles, and the second part of the sum adapting that logic for primary vehicles, who will only be contribute the patient's survival if all faster secondary vehicles are also busy. These interpretations are given as annotations to the equations in Appendix A.

4.2. A Cross-Entropy Optimisation Approach

Cross-entropy was first proposed as a method for finding the optimal solution of combinatorial and continuous non-convex optimisation problems with convex bounded domains [31]. It has since been applied to a broad range of problem types [10]. Here a population-based heuristic is implemented to find allocations that maximise the MESLMHPHF score, using cross-entropy. The full implementation is given at `https://github.com/MarkTuson/ej_hphf`. The central idea behind this approach is a matrix of probabilities, with entries p_{az}, representing the probability that z is the optimal number of ambulances required at ambulance station a.

The heuristic progresses iteratively, with each iteration generation a new population of ambulance allocations by sampling from the probability matrix. That population is ranked according to the MESLMHPHF score, and a

proportion of the best performing allocations are used to update the p_{az} values according to a smoothing factor. This is repeated until no improvement is found in the score of the top-ranking allocation between iterations.

5. Simulating Allocations

Discrete event simulation (DES) is a common methodology that allows us to investigate given scenarios under uncertainty. Here it will be used to quantify the effectiveness of given ambulance allocations by measuring a range of key performance indicators (KPIs), such as the average response time, the percentage of abandoned calls, vehicle utilisations, and expected survival based on response times. The simulation is built using the Ciw library [26], and the model logic is described below. Its central ideas include modelling transit jobs as customers, rather than the patients themselves, and simulating primary and secondary vehicles as two simulations sequentially, with the output of the former being the input of the latter. First, Section 5.1 introduces some notation used throughout the paper.

5.1. Notation

Let $\mathscr{P}$ be the set of pick-up locations, indexed by p; $\mathscr{A}$ be the set of ambulance locations, indexed by a; $\mathscr{Y}$ be the set of hospitals, indexed by y; and $\mathscr{K}$ be the set of medical specialities, indexed by k. An allocation is given by Z_a and $\tilde{Z}_a$, where Z_a is the number of primary vehicles allocated to ambulance location a; and $\tilde{Z}_a$ is the number of secondary vehicles allocated to ambulance location a.

Now we also let:

- $\tilde{B}_{pa}$ be the traffic-free travel time from ambulance location a to pick-up location p; and let B_{pa} be a random variable representing the time it takes for an ambulance to drive this distance;
- $\tilde{C}_{py}$ be the traffic-free travel time from pick-up location p to hospital y; and let C_{py} be a random variable representing the time it takes for an ambulance to drive this distance;
- $\tilde{D}_{ya}$ be the traffic-free travel time from hospital y to ambulance location a; and let D_{ya} be a random variable representing the time it takes for an ambulance to drive this distance;

- $\tilde{F}_{pa}$ be the traffic-free travel time from pick-up location p to ambulance location a; and let F_{pa} be a random variable representing the time it takes for an ambulance to drive this distance;
- G_k be a random variable representing the time the ambulance spends with patients of speciality k at the pick up location;
- J_k be a random variable representing the time the ambulance spends with patients of speciality k at the hospital;
- Θ be a random variable representing the time the ambulance spends re-fuelling, re-stocking, and resting between transit jobs;
- λ_{pk} be the rate at which patients of speciality k make calls from pick-up location p;
- q_{pky} be the probability that a patient of speciality k from pick-up location p is taken to hospital y. Note that $\sum_y q_{pky} < 1$ is possible, that is a patient may not go to any hospital, in which case the ambulance returns to their ambulance location.

5.2. Simulating Primary Vehicles

For primary vehicles, that is emergency ambulances, the logic to simulate is as follows: Patients make a call from one of a number of pick-up locations and await an ambulance to pick them up and take them to an appropriate hospital. Ambulances are stationed at a number of ambulance locations; when a patient makes a call, all free ambulances from any location calculate their expected time to reach that patient, and the ambulance with the smallest expected time to the patient is called out. The ambulance drives from its current ambulance location to the patient's pick-up location, then if a hospital is required, from the patient's pick-up location to the hospital, and then from the nearest hospital back to their original ambulance location. If a hospital is not required, the ambulance returns to their original ambulance location.

Rather than considering patients as customers in a queue, here the situation is re-framed to model transit jobs as customers, with the ambulances as servers.

Transit jobs can be categorised into classes corresponding to the pick-up locations $\mathscr{P}$ and speciality $\mathscr{K}$. Jobs of class $(p,k) \in \mathscr{P} \times \mathscr{K}$ arrive with rate

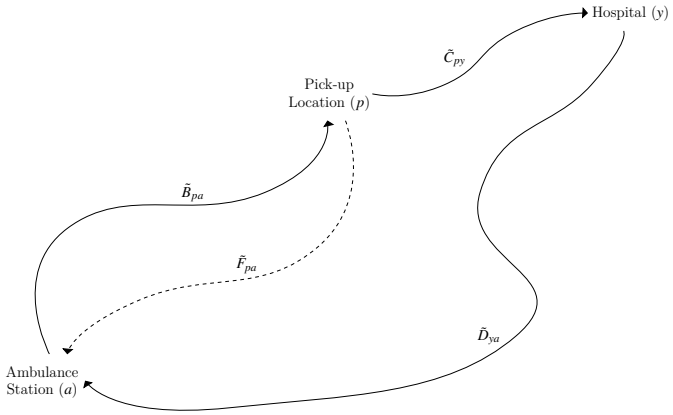


Figure 2: Ambulance routes for a given transit job.

λ_{pk}. Servers can be categorised into classes corresponding to the ambulance locations $\mathscr{A}$. Service times are server-dependent, that is the service time of a transit job of type (p,k), being served by a server of class a, will have service time T_{paky} given by Equation 10 if a hospital is required, and service time T_{pak} given by Equation 11 if no hospital is required.

$$T_{paky} = B_{pa} + G_k + C_{py} + J_{yk} + D_{ya} + \Theta \tag{10}$$

$$T_{pak} = B_{pa} + G_k + F_{pa} + \Theta \tag{11}$$

That is T_{paky} and T_{pak} is the overall time the ambulance spends dealing with that transit job and is unavailable to receive any more transit jobs. Figure 2 shows the route a particular ambulance will take when servicing a transit job, where there is a probability $1-\sum_y q_{pky}$ that the ambulance will return, possibly on a different route, assuming distances are not symmetrical, to the ambulance location after a pick-up (dashed line), when patients do not need to go to the hospital.

From a patient's point of view their service only lasts $G_k + C_{py} + J_{yk}$, (or G_k if no hospital is required), and their waiting time is B_{pa} plus the time waiting for an ambulance to be dispatched. However in our case, if there are no free ambulances at the time of call, it is assumed that patients find their own care or transport, and so the call is abandoned. Therefore, the time waiting for dispatch is always zero.

An ambulance from location $a_\star$, from the set of currently free ambulances $\mathscr{A}_{\text{free}}$, is assigned to a call from pick up location p by:

$$a_\star = \arg\min_{a \in \mathscr{A}_{\text{free}}} \mathbb{E}[B_{pa}]. \tag{12}$$

Furthermore, all service times are time dependent, and calculated from travel distances and approximate hourly traffic levels, described in Section 5.3.

5.3. Travel Time Calculations

Each day is split into a set of periods, $\mathscr{H}$ indexed by h. Within each period traffic levels are considered constant but differ from period to period. Traffic levels influence the speed of the ambulance through a delay factor d_h associated with each period. Therefore, the expected speed S_h at which an ambulance travels during period h is given by Equation 13,

$$S_h = s d_h \tag{13}$$

where s is the average speed that ambulances drive given no traffic. The relationship between time travelled and distance covered is piece-wise linear with slope S_h in period h. Therefore, travel times are calculated using this relationship.

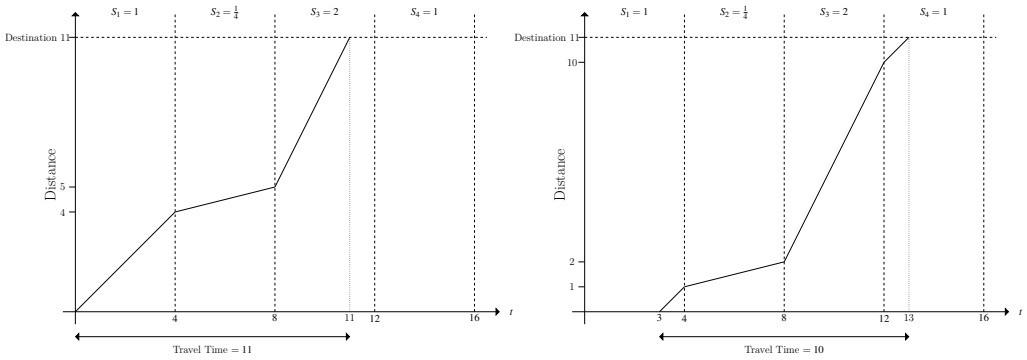
As an example, consider the scenario where for the first period $t \in (0,4)$ we have $S_1 = 1$; for $t \in (4,8)$ we have $S_2 = \frac{1}{4}$, for $t \in (8,12)$ we have $S_3 = 2$, and for $t \in (12,16)$ we have $S_4 = 1$. Consider that the ambulance must travel 11 units. If the ambulance begins its journey at $t = 0$, then the expected travel time would be 11 time units, shown in Figure 3a. However if the ambulance begins its journey at time $t = 3$ then the expected travel time is 10 time units, shown in Figure 3b.

Furthermore it is assumed that travel times for each segment of the route follow a Triangular distribution around the calculated expected travel time, between 75% and 125% of the value. This is the method used to calculate the expected values of B_{pa}, C_{py}, D_{ya} and F_{pa} for the transit job service times, from the traffic-free travel times $\tilde{B}_{pa}$, $\tilde{C}_{py}$, $\tilde{D}_{ya}$ and $\tilde{F}_{pa}$ respectively.

5.4. Time-dependent Demand

Additionally, the call arrival rates λ_{pk} are Poisson and time dependent. Each day is split into four 6-hour periods, the morning, afternoon, evening and night, and each speciality k and pick up location p will have a different call arrival rate for each of these periods.

For some locations and specialities the number of observed calls might be very low, and so the λ_{pk} rate would be very low. This can cause synchronicity



(a) Beginning at $t = 0$, a travel time of 11.

(b) Beginning at $t = 3$, a travel time of 10.

Figure 3: Example of calculating travel times for two different starting times.

issues when sampling arrivals [27], where artificially long inter-arrival times can be introduced at the beginning of one time period due to the low arrival rate in the previous time period. In order to overcome this here, rather than sample inter-arrival times iteratively from an Exponential distribution, an entire schedule of arrival dates are sampled at the beginning of the simulation run, first by sampling the number of arrivals required in each time period from a Poisson distribution, and then by sampling specific dates within that time period using a Uniform distribution.

This mechanism was implemented in the Ciw software in version v2.3.3 [33], and described in the documentation here: `https://ciw.readthedocs.io/en/latest/Guides/time_dependent.html`.

5.5. Simulating Secondary Vehicles

Secondary non-transit vehicles, such as rapid response vehicles (RRVs), can travel faster than primary transport vehicles such as ambulances, and so can be dispatched at the same time as the primary vehicle but reach the patient earlier, providing vital on-site care before the primary transit vehicle arrives. Here secondary vehicles are simulated in a second discrete event simulation, run sequentially, but simulating the exact same time period and events as the first simulation. This is similar to sequential hybrid simulation methodology [7, 25], however in this case the two combined components are both DES, and are combined to simplify the logic of each component.

This is possible as there are no synchronicity issues between the two components. Primary vehicles operate independently of secondary vehicles, that is the way primary vehicle responds to a patient is unaffected by the presence or lack of a secondary vehicle. Therefore, primary vehicle logic is not compromised by simulating primary vehicles in isolation. Secondary vehicles are impacted by the behaviour of primary vehicles, they must remain with the patient until the primary vehicle arrives, and so must be simulated after the simulation of primary vehicles, taking as inputs the exact list of events that occurred. That is the logic of the secondary vehicles is determined by observing the actions of primary vehicles and reacting to them. Simulation results are combined for each individual patient, to determine the their response time caused by either primary or secondary vehicles.

Secondary vehicles are chosen in the same way was primary vehicles, by choosing out of the currently free vehicles the one with the smallest expected time to patient, given in Equation 12. Their service times are reactionary to what occurred with the primary vehicle.

Let B_{pa_2} be the time it takes for the secondary vehicle to travel from its location a_2 to the patient pick-up location p; $\hat{B}_{pa_1}$ is the time is took for the primary vehicle to get from its location a_1 to the patient pick-up location p; $\hat{G}_k$ is the time the primary vehicle spent with the patient; and F_{pa_2} is the time is takes to return to the secondary vehicle's location a_2 from the patient pick-up location p. Note that $\hat{B}_{pa_1}$ and $\hat{G}_k$ are exact values obtained from the initial simulation of the primary vehicles, while B_{pa_2} and F_{pa_2} are random variables to sample in the subsequent simulation. Travel times are calculated as described in Section 5.3, replacing the primary vehicle delay factor, d_h, with a delay factor for secondary vehicles, $\tilde{d}_h$. This accounts for secondary vehicles travelling faster and reacting to traffic differently to primary vehicles.

Exact synchronicity considerations are shown in Figure 4:

- whenever the secondary vehicle reaches the patient before the primary vehicle, they remain with the patient until the primary vehicle leaves;
- if secondary vehicle arrives after the primary vehicle they will immediately return to their stations as they are not needed at the scene;
- if the expected time for the secondary vehicle to reach the patient exceeds the expected time for the primary vehicle to reach the patient then the secondary vehicle is not deployed, and so that transit job

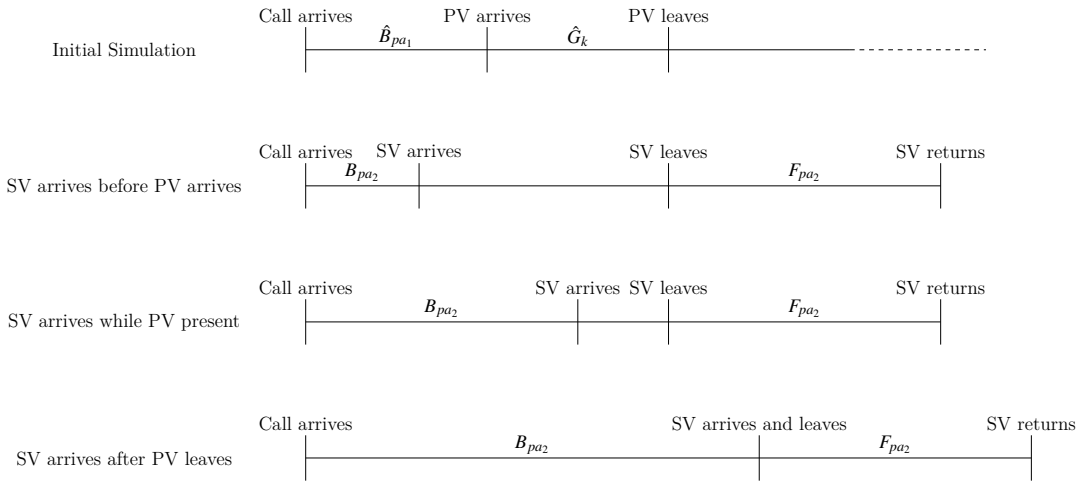


Figure 4: Visualising secondary vehicle logic reacting to the primary vehicle's actions.

would be abandoned in the secondary simulation (although still seen by a primary vehicle in the initial simulation).

Therefore job service times for secondary vehicles, $\tilde{T}_{pak}$, are given by:

$$\tilde{T}_{pak} = \max\left(B_{pa_2}, \hat{B}_{pa_1} + \hat{G}_k\right) + F_{pa_2} \tag{14}$$

5.6. Combined Model

Combining the logic presented in Sections 5.2 and 5.5 gives the overall simulation logic for simulating both primary and secondary vehicle types. It is used to quantify the effectiveness of a given allocation $(Z_a, \tilde{Z}_a)$ for all $a \in \mathscr{A}$, by recording several useful KPIs. In this work, the KPIs of interest are: the average primary vehicle utilisation, the average secondary vehicle utilisation, the mean response time, the percentage of abandoned calls (that is a measure of the primary vehicle unavailability), and the expected overall survival. Survival is modelled using a combination of survival function curves and hard cut-offs, and is described in Section 3.

6. The Case of Jakarta

Jakarta, the capital of Indonesia, has a population of 10.5 million [18] residing within 664 km^2. The city is divided into five municipalities and one

district, each of which is divided into sub districts and, in turn, neighbourhoods. There is a total of 42 sub districts and 261 neighbourhoods within mainland Jakarta (excluding the Thousand Islands regency in the north) [17], a map is given in Figure 5.

As described in Section 1, there are many challenges to calling for an ambulance in Jakarta, as captured in [8]. A fragmented ambulance service and poor data collection also means that estimating the true demand for ambulances in Jakarta is itself particularly challenging. To help understand these difficulties, we ran workshops and worked with, and collected data from, different ambulance providers, including Ambulans 118 and subsequently the 119 Emergency Ambulance Service, which was recently established by the Indonesian Government as an attempt for a coordinated EMS system. Based on this survey work [8] and initial analysis and reporting, the regional government of Jakarta invested in a new fleet of coordinated ambulances accessible via the single emergency number 119. This work evaluates this fleet's effectiveness in terms of a range of KPIs, including survival.

The developed models have been parameterised using demand data that covers all 261 neighbourhoods from 1 January to 31 December 2019, before the COVID-19 pandemic. 2020-2021 data received was, naturally, heavily skewed by the pandemic with significantly lower demand, and so was not considered as representative of a typical period for forecasting future needs, hence the decision to use the available 2019 demand.

Working with our ambulance partners, after careful consideration the following patient 'specialities', that is priority categories, were agreed and have been used in our work to prioritise demand based on clinical need (although other categories could readily be adopted and included as necessary):

- A1 - High priority emergency patients: critical patients who require immediate life-saving assistance with a target ambulance response time of 8 minutes. In our data set there were 168 identified calls that met this criterion.

- A2 - Other emergency patients: urgent patients that require assistance with a target response time of 15 minutes. In our data set there were 23,784 calls of this type.

- B - Non-emergency patients: patients that still benefit from an ambulance response but non-critical with a target response time of 60 minutes. There were 31,725 calls of this type in the data set.

Figure 5: Map of Jakarta's Municipalities.

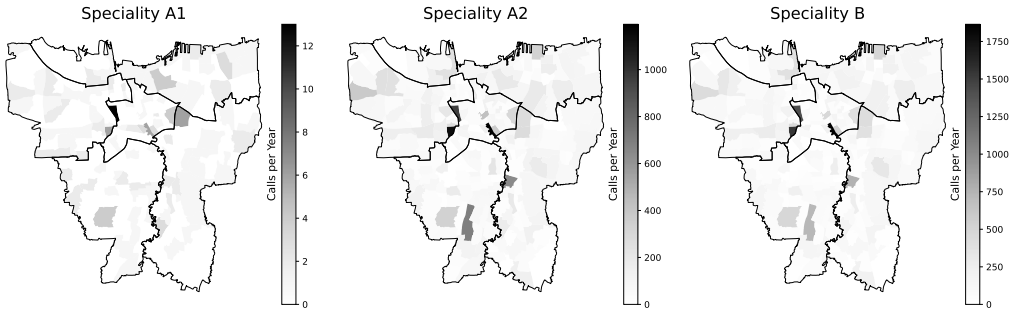


Figure 6: Number of calls received by speciality and neighbourhood.

Municipality	Number of neighbourhoods	Current number of ambulance stations
North	31	9
South	65	14
East	63	13
West	57	16
Central	45	15

Table 2: Numbers of neighbourhoods and ambulance stations by municipality.

The number of calls from each of these specialities varies by neighbourhood, as shown in Figure 6. It can be seen that many calls are highly concentrated in a handful of neighbourhoods rather than spread throughout the city.

In Sections 6.1 and 6.2 we use the discrete event simulation described above to find appropriate KPIs to measure the performance of the current allocations and two proposed grid allocations for the city. Appendix B provides further details on the parameterisation of the model.

6.1. Current Allocation

As of October 2022, the 119 service in Jakarta ran a fleet of 81 primary Emergency Ambulances (EAs) and 13 secondary motorbike Rapid Response Vehicles (RRVs), distributed across 67 ambulance stations throughout the city. Table 2 and Figure 7 summarise the allocations by municipality and sub districts.

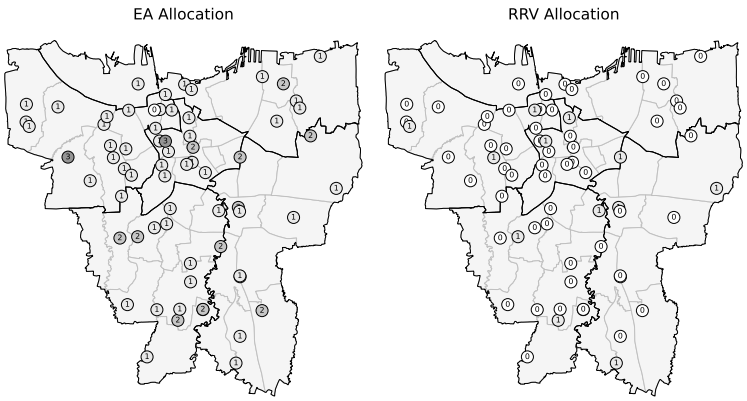


Figure 7: Map of current Emergency Ambulance (EA) and Rapid Response Vehicle (RRV) locations and allocations to ambulance bases across Jakarta.

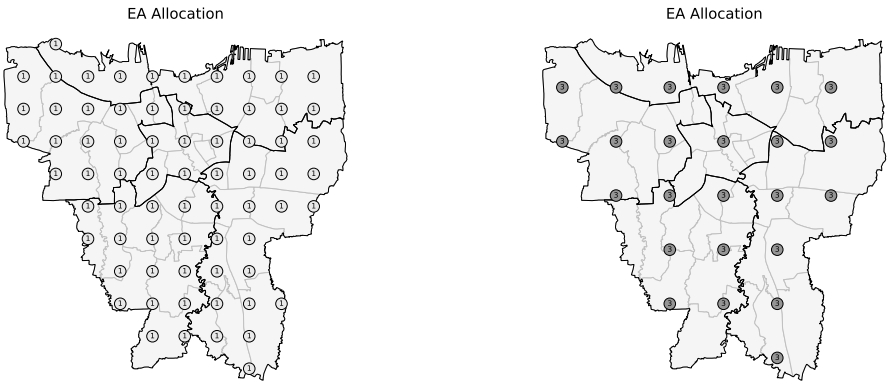
6.2. Proposed 'Grid' Allocation

Discussions with senior staff at 119 identified that they were seriously considering to re-configure their ambulance allocations into a grid structure, placing an ambulance station at regular intervals throughout the city and uniformly distributing the vehicles that they felt would ensure coverage. Two possible grid allocations were being considered: one placing an EA every 3km across the city (giving a total of 70 vehicles), and another placing three EAs every 5km across the city (giving a total of 72 vehicles). Figure 8 show these proposed allocations.

The simulation was run for both current and proposed grid allocations and results are shown and compared in Table 3. We observe that the two proposed grid allocations perform much worse than the current allocation as measured by mean response time, survival, and percentage of calls abandoned.

6.3. Planning for Future Demand

Similarly to many other cities worldwide, ambulance services in Jakarta anticipate that demand for services will grow in future years, not least that there is now a coordinated service accessible via a single common number



(a) 3km grid.

(b) 5km grid.

Figure 8: Proposed grid allocations.

Allocation	Baseline	Grid 3km	Grid 5km
Number of EAs	81	70	71
Number of RRVs	13	0	0
Ambulance Utilisation	31.7%	37.4%	36.6%
RRV Utilisation	40.9%	-	-
Mean Response Time (mins)	16.80	22.39	24.14
Overall Survival	75.5%	63.7%	61.0%
Percent Abandoned	0.034%	0.430%	1.028%

Table 3: Calculated KPIs for the current and proposed grid allocations.

Reason	% of emergency respondents
Used Ambulance	13.0
Too expensive	5.2
Not available	13.1
Would take too long	16.5
Not necessary	15.7
Not aware of service	33.1
Other	3.5

Table 4: Barriers to use of an ambulance service in Jakarta, from [8].

'119' to call. This should help raise awareness and visibility of an ambulance service. In the case of Jakarta, certainly increased use of 119 and therefore increased demand would be a desired outcome resulting from proactive actions by the Indonesian Government.

We consider four different demand scenarios, developed by considering the result of our cross-sectional study of patients attending EDs in Jakarta [8]. As part of the survey, randomly selected patients arriving at each emergency department were asked whether they had used an ambulance to attend, and reasons for not using an ambulance. For those patients who were categorised by the medical staff as emergency and for whom therefore an ambulance might have seemed a sensible option, the responses are given in Table 4.

The survey confirmed three major barriers to use, namely: the service's visibility (33.1% of respondents were unaware of the service), the service's reliability (13.1% of respondents reported that there was no ambulance available and 16.5% of respondents believed it would take too long), and the service's cost (5.2% of respondents cited cost as a deterrent). Therefore, four demand scenarios were considered, corresponding to addressing each of these barriers in turn:

- demand_13: this represents the current situation where approximately 13% of emergency patients (specialities A1 and A2) do use an ambulance.

- demand_19: this represents the situation where visibility is addressed. Here, we distribute the 33.1% of the respondents who were unaware of the service proportionally between using the ambulance and amongst

Demand Scenario	demand_13	demand_19	demand_34	demand_45
Ambulance Utilisation	31.7%	38.9%	54.9%	62.5%
RRV Utilisation	40.9%	47.7%	62.3%	68.5%
Mean Response Time (mins)	16.80	17.98	22.18	23.97
Overall Survival	75.5%	67.7%	52.6%	46.0%
Percent Abandoned	0.034%	0.315%	3.910%	10.004%

Table 5: Calculated KPIs for the current allocation under the four possible demand scenarios.

the remaining issues. Using this methodology we would expect 19.4% of emergency patients to now use an ambulance.

- demand_34: this represents the situation where reliability and visibility are addressed. Using the same methodology we would expect 34.8% of emergency patients to now use an ambulance.

- demand_45: this represents the situation where cost, reliability and visibility are all addressed. Using the same methodology we would expect 45.4% of emergency patients to now use an ambulance.

Guided by our findings in [8], recent changes mean that the 119 services is now free to use for Jakarta residents, so certainly the highest demand (demand_45) scenario is plausible in the near future. After discussions with our partners, it was agreed that for each of the scenarios we should assume that non-emergency demand (speciality B) remains unchanged.

Table 5 gives the derived KPIs for the current allocation for each of these demand scenarios in the simulation. In all measured KPIs, the performance worsens as emergency demand increases, thus motivating the need to find better allocations and investigate the effect of varying resource levels and explict consideration of patient outcomes, detailed in Section 4.

6.4. Improving the Current Allocation

The results here are not real, this is a placeholder section for when we get the results from Mark.

Applying the cross-entropy heuristic to the current allocation re-allocates the 81 EAs and 13 RRVs. Figure 9 shows the new allocation (in comparison

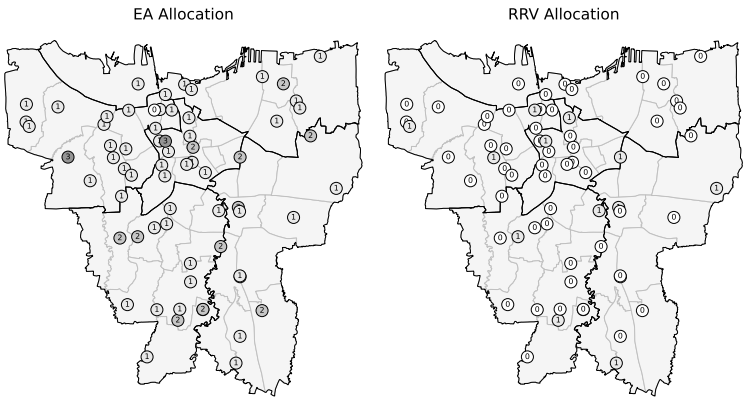


Figure 9: Improved allocation of 81 EAs and 13 RRVs.

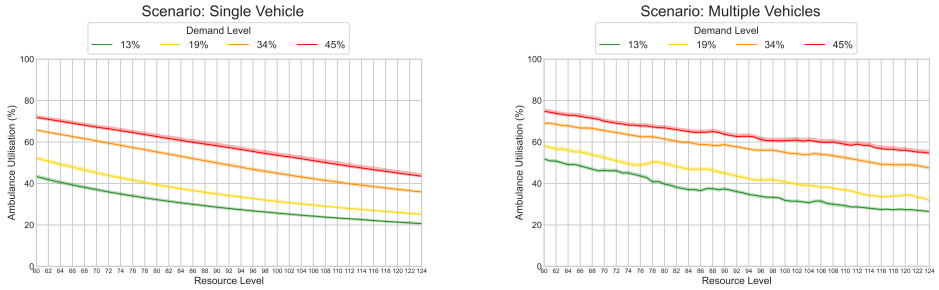
to Figure 7. Table 6 compares the simulation derived KPIs between the current and the optimised allocations; it can be seen that a number of them increase.

6.5. Experimental Design

The heuristic algorithm described in Section 4.2 finds allocation of ambulances across the 67 current ambulance stations for a given resource level, that is for a given number of EAs and RRVs. We perform experiments to find allocations for each demand scenario described in Section 6.3, for resource levels between 60 and 124 EAs, and for two scenarios: a single vehicle

Allocation	Baseline	Improved
Ambulance Utilisation	31.7%	31.7%
RRV Utilisation	40.9%	40.9%
Mean Response Time (mins)	16.80	16.80
Overall Survival	75.5%	75.5%
Percent Abandoned	0.034%	0.034%

Table 6: Comparing KPIs for the current and improved allocations.



(a) Single vehicle type scenario.

(b) Multiple vehicle type scenario.

Figure 10: Ambulance utilisation results.

type scenario with no RRVs, and a multiple vehicle type scenario with both EAs and RRVs. Information supplied by one of the ambulance operators suggested that in terms of total running costs one EA was approximately equivalent to three RRVs, and so we consider three RRVs to be one resource.

Run times of the cross-entropy heuristic for allocating all 67 ambulance stations proved to be impractical. Therefore, working on the assumption that the best allocations within each municipality were independent of one another, the heuristic was run on each of the five municipalities (Table 2) independently and the allocations combined. Initially each municipality was allocated a resource level proportional to their population, and as the experiments increased the overall resource level, this was added to the municipality with the worst resource-population ratio, and then the heuristic was re-ran with it's new resource level.

6.6. Optimisation Results

Allocations were found for each resource level between 60 and 124 EA equivalents, for both single and multiple vehicle scenarios, for each of the four demand scenarios. Figures 10-14 display the obtained KPIs for each of these allocations. It can immediately be seen that increasing the resource level has a positive effect on all KPIs, with vehicle utilisations, mean response times, and percentage abandoned decreasing, and overall survival increasing. Similarly, as expected increasing the demand of emergency calls has a negative impact on all KPIs.

It is particularly interesting to compare the scenario in which single vehicle types (only emergency ambulances) were allocated against the scenario

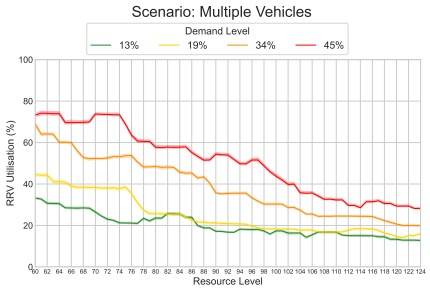
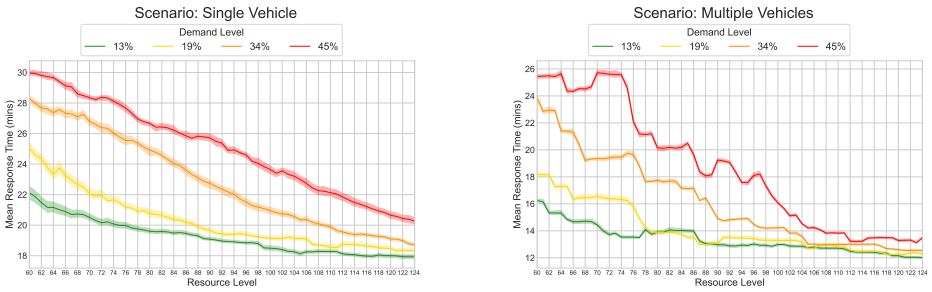


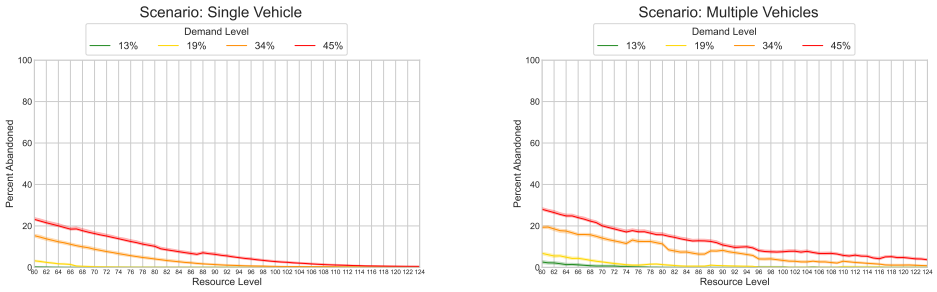
Figure 11: RRV utilisation results.



(a) Single vehicle type scenario.

(b) Multiple vehicle type scenario.

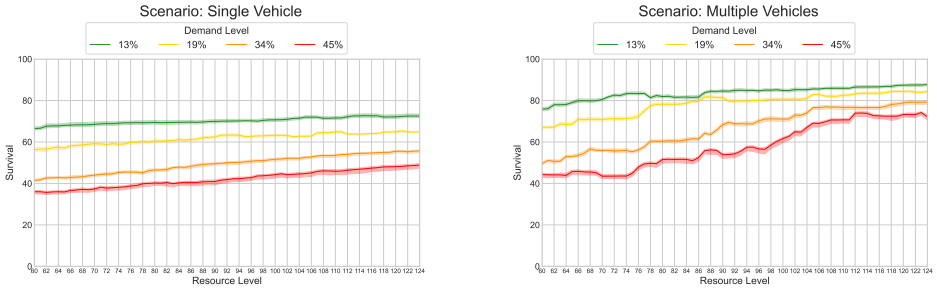
Figure 12: Mean response time results.



(a) Single vehicle type scenario.

(b) Multiple vehicle type scenario.

Figure 13: Percent of abandoned calls results.



(a) Single vehicle type scenario.

(b) Multiple vehicle type scenario.

Figure 14: Overall survival results.

where multiple vehicle types (both emergency ambulances and rapid response vehicles) were allocated. For equivalent resource levels, introducing secondary vehicles increases ambulance utilisation, and so decreases availability, resulting in an increase in abandoned calls. However, introducing secondary vehicles also results in a large decrease in the mean response times, and so a large increase in the percentage of patients seen within target.

It is also noticeable that the increases or decreases in the simulation derived KPIs are not necessarily monotonic with increases in resource level. This is due to the derived allocations being based on the MESLMHPHF score only, which is a measure of survival, while the simulation derived KPIs are indirect measures of performance. This may indicate that these indirect measures of performance, such as mean response times or percentage within target, are not good indicators of survival.

Crucially, the plots show that the allocations produced by the optimisation algorithm perform better than the current allocation for the same demand and resource levels. The current levels include 81 EAs and 13 RRVs, so a resource level of 85.33, with multiple vehicles. Comparing the derived allocations for this level in Figures 10-14 to the results given in Table 3, the derived allocations give lower ambulance and RRV utilisations, lower mean response times, and a higher percentage of patients seen within target.

As experiments were run across all demand scenarios, we can see how much more resources would be required, if optimally located, to achieve a similar level of service as is currently offered, when demand increases. Table 7 shows, for each demand scenario, the minimum resource level and vehicle type

Demand Scenario	Mean Response Time	Overall Survival	Percent Abandoned
demand_13	60, (50, 30)	60, (50, 30)	80, (66, 42)
demand_19	70, (60, 30)	78, (63, 45)	101, (79, 66)
demand_34	88, (75, 39)	106, (83, 69)	> 124
demand_45	99, (85, 42)	> 124	> 124
Current Level	16.8 mins	75.48%	0.0339%

Table 7: Resources required to maintain current KPI levels when optimising locations under the four demand scenarios.

breakdown required to meet the current level in each of the following KPIs: mean response time, overall survival, and percentage of calls abandoned. For example, in order to maintain the current 75.48% overall survival, under the demand_19 scenario the city would require 63 EAs and 45 RRVs. It is noticeable that fewer resources are required to maintain the current mean response time than to maintain the current overall survival.

7. Discussion & Conclusions

This paper describes the development of EMS demand and capacity models and their application to the city of Jakarta, Indonesia. To our knowledge, this is the first such study of analysing emergency ambulance services in Jakarta, and the work has directly assisted and guided providers and the government to make investment decisions for a coordinated and free to use EMS.

Firstly, a discrete event simulation model has been used to evaluate existing and potential heterogeneous vehicle ambulance feet allocations in terms of key performance measures, such as response times, response time targets, and vehicle utilisation rates. This takes into consideration geospatial demand and travel, and temporal variation in demand and traffic levels. A novel feature of our approach is that the model comprises of sequential simulations, feeding data directly from one into the other in order to simulate primary and secondary vehicles separately while maintaining synchronicity, with the overall aim of reducing the complexity of the simulation logic.

Secondly, we have considered patient survival and outcomes within a developed cross-entropy heuristic. By considering geospatial travel times and

accounting for the varying needs of different patient types though categorising patients by priority and through using different survival function profiles, an expected overall survival was given. This could be used within a optimisation processes, in this case a cross-entropy heuristic, to find vehicle fleet allocations that maximise expected survival. This expected survival function extends previous work in [19] to include more than one vehicle type.

Both models investigate ambulance service activity in the case of heterogeneous patients and heterogeneous fleets. Heterogeneous patient groups consider those with distinct demand profiles, priorities, and survival functions. Heterogeneous fleets concern different types of vehciles, in our case emergency ambulances (EAs) which respond to every patient, and Rapid Response Vehicles (RRVs) which can be utilised to reach patients faster, despite being unable to transport patients themselves.

Using a combination of models has permitted an approach that can capture performance measures or take into account factors that each model in isolation can not. For example, the maximum expected survival model is a lot more appropriate for use within an optimisation process as the runtimes are a far more reasonable than the simulation. On the other hand, the simulation allows for greater complexity such as temporal demand, and is able to capture a greater range of KPIs.

A key feature of both models, and their novelty, is the consideration of heterogeneous fleets. In a highly populated area such as Jakarta (around 16,000 population per km^2), the inclusion of RRVs such as paramedics on motorbikes is very important, given RRVs can access areas that cannot easily or quickly be accessed by ambulances. Results from both the optimisation and simulation research showed that the RRVs can help increase the overall survival and reduce the response times, especially crucial when responding to life threatening events such as cardiac arrests [14].

The results of our research also suggest that allocation strategies may not be intuitive to ambulance service managers, further emphasising the value of a modelling approach. For example, senior managers suggested the use of a grid allocation to maximise geographic coverage, with ambulances equally spread across the city. This was shown to be sub-optimal because neighbourhoods have different population densities and characteristics. For example, the number of daily commuters in Jakarta is considerably high during day time due to work and education [16]. Municipalities where trading, educational institutions, and offices are concentrated may become more populated during day time. The model has quantified the deterioration in

response times and patient outcomes should a fixed grid system be implemented, leading to 119 to drop this consideration.

The ambulance posts in Jakarta depend on the service provider. Those ambulances provided by the government are in various locations including government buildings as well as community clinics and sub-district hospitals. In our case, we used current ambulance posts as the locations for allocating the resources. As demands for ambulances may change from time to time, future work could evaluate different (and perhaps dynamic) ambulance posts that depend on changing demand volumes in neighbourhoods by time of the day.

The data used for ambulance demand covered one year from 1 January to 31 December 2019, prior the COVID-19 pandemic. Recent studies related to ambulance demand indicate that the pandemic has severely impacted on the utilisation of EMS, for examples in call volumes [32] and in specific medical conditions such as trauma [3], and possibly still continue to affect demand. Ambulance providers, such as 118 and 119, may use our modelling tools to incorporate future demand data and re-evaluate resource needs. To aid this, we are currently working on interfacing the simulation and optimisation into a single easy to use decision support tool with a dashboard (data visualiser).

Studies have shown that the quality in pre-hospital data collection varies considerably in Indonesia [15]. We encountered similar challenges and have suggested to 119 and the Indonesian Government that they should continue to strive to improve the coverage and quality of data collection. Future work could explore in more depth unmet demand that is not recorded in the ED or in the ambulance data.

Indonesia is a vast country with an uneven population density and varying quality of available healthcare resources. The models have been parameterised with data from a high population, high density urban area which is also relatively well resourced with healthcare facilities compared to other regions in the country. Ambulance services in rural areas of Indonesia may not have the same quality in management and organisation compared to the capital, Jakarta. We intend to conduct future studies to apply the developed models to analyse ambulance demand and allocations in different areas of Indonesia, including in rural regions.

In conclusion, the developed models have demonstrated a novel approach in modelling ambulance allocations that incorporate multiple vehicle types and health conditions, whilst also capturing patient survival. Our work has already informed major decisions on the design of a free to use and coordi-

nated EMS system for Jakarta. Ongoing collaboration will continue to assist ambulance providers and the Indonesian Government in providing evidence-based decision making for the benefit of patients and the population they serve, including the exploration of the roll-out of 119 beyond Jakarta to other regions of Indonesia.

CRediT authorship contribution statement

From Computers and OR, select from:

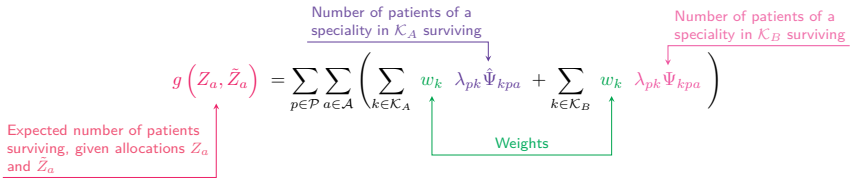
Conceptualisation; Funding acquisition; Investigation; Methodology; Project administration; Resources; Software; Supervision; Validation; Writing - original draft; Writing - review & editing.

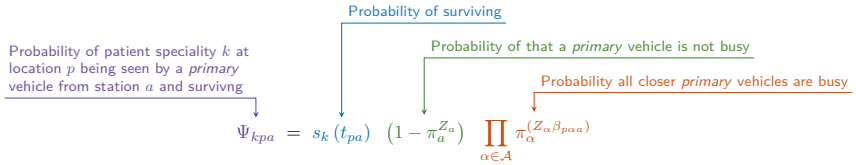
Geraint Palmer: Conceptualisation, Investigation, Methodology, Software, Writing – original draft
Sarie Brice: Conceptualisation, Investigation, Methodology, Writing – original draft
Daniel Gartner: Conceptualisation, Funding acquisition, Methodology, review & editing.
Paul Harper: Conceptualisation, Investigation, Methodology, Funding acquisition, Resources, Project administration, Writing – review & editing.
Vince Knight: Conceptualisation, Funding acquisition, Methodology, Software, review & editing.
Leanne Smith: Methodology.
Mark Tuson: Conceptualisation, Investigation, Methodology, Software, Writing – original draft

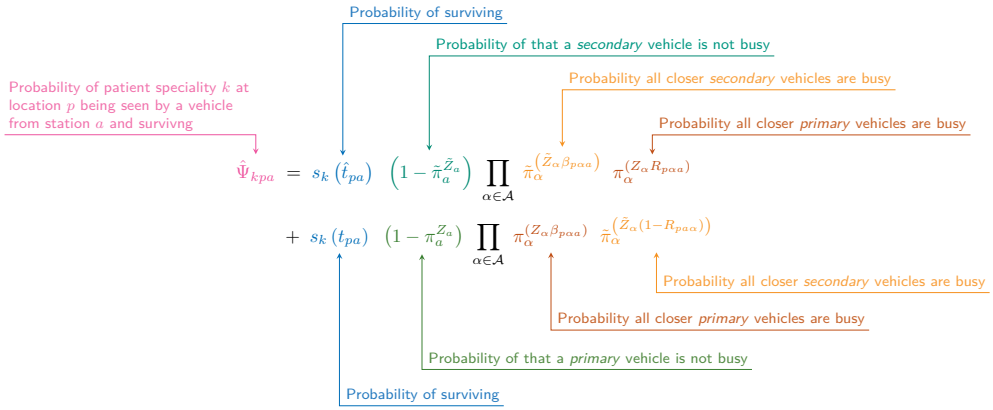
Acknowledgements

The study was funded by EPSRC with grant no: EP/T003197/1. We would like to express our sincere gratitude to Indonesian emergency ambulance providers including 118 and 119 for their support, insights and provision of data. In particular we wish to thank Professor Aryono Djuned Pusponegoro and Ms Asti Puspita Rini, Founder and Director of 118 Ambulance Service Foundation and Dr Winarto, Head of the 119 Ambulance Service in Jakarta.

Appendix A. Annotated Objective Function







Appendix B. Model Parameters

Call arrival rates λ_{pk} are derived from 2019 demand data slit by municipality and speciality, and time of day. Probabilities q_{pky} are similarly derived from the 2019 data of transit journeys. All traffic-free travel times are found using Google Maps API, while time-dependent traffic delays are found from the TomTom website. The other models are parameterised in the following way:

- From a sample of calls the time at site G_k was found to follow a lognormal distribution with parameters $\mu = -0.6219$ and $\sigma = 0.8048$. For the case of Jakarta this was modelled identically for all specialities k. Comparison between the lognormal fit and the sampled times are given in Figure B.15.

- From discussions with staff at the ambulance service in Jakarta, the time at hospital J_k was modelled as a Uniform distribution between 40 and 60 minutes for the emergency specialities A1 and A2, and between 20 and 30 minutes for non-emergency speciality B.

- From discussions with staff at the ambulance service in Jakarta, the refill time Θ is taken to be 60 minutes for an emergency ambulance, and 15 minutes for an RRV.

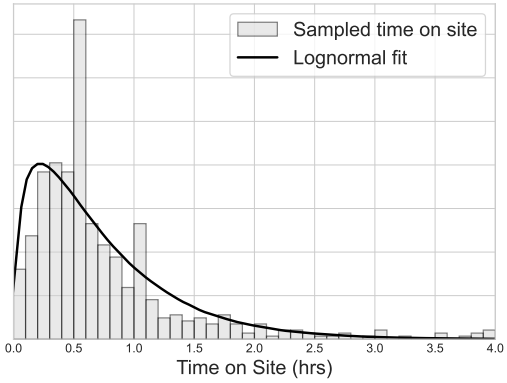


Figure B.15: Comparison between the sampled time on site and the lognormal fit.

References

[1] Marco Amorim, Francisco Antunes, Sara Ferreira, and António Couto. An integrated approach for strategic and tactical decisions for the emergency medical service: Exploring optimization and metamodel-based simulation for vehicle location. Computers & Industrial Engineering, 137:106057, 2019.

[2] Roberto Aringhieri, Maria Elena Bruni, Sara Khodaparasti, and J Theresia van Essen. Emergency medical services and beyond: Addressing new challenges through a wide literature review. Computers & Operations Research, 78:349–368, 2017.

[3] Michael Azbel, Mikko Heinänen, Mitja Lääperi, and Markku Kuisma. Effects of the covid-19 pandemic on trauma-related emergency medical service calls: a retrospective cohort study. BMC emergency medicine, 21(1):1–10, 2021.

[4] Valérie Bélanger, A Ruiz, and Patrick Soriano. Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles. European Journal of Operational Research, 272(1): 1–23, 2019.

[5] Lauren E Birmingham, Andrea Arens, Nyaradzo Longinaker, and Colleen Kummet. Trends in ambulance transports and costs among medicare beneficiaries, 2007–2018. The American Journal of Emergency Medicine, 47:205–212, 2021.

[6] Justin J Boutilier and Timothy CY Chan. Drone network design for cardiac arrest response. Manufacturing & Service Operations Management, 24(5):2407–2424, 2022.

[7] Sally C Brailsford, Tillal Eldabi, Martin Kunc, Navonil Mustafee, and Andres F Osorio. Hybrid simulation modelling in operational research: A state-of-the-art review. European Journal of Operational Research, 278(3):721–737, 2019.

[8] Syaribah Noor Brice, Justin J Boutilier, Daniel Gartner, Paul Harper, Vincent Knight, Jen Lloyd, Aryono Djuned Pusponegoro, Asti Puspita Rini, Jonathan Turnbull-Ross, and Mark Tuson. Emergency services utilization in Jakarta (Indonesia): a cross-sectional study of patients attending hospital emergency departments. BMC health services research, 22(1):639–639, 2022. ISSN 1472-6963.

[9] CY Chang, S Abujaber, TA Reynolds, CA Jr Camargo, and Obermeyer Z. Burden of emergency conditions and emergency care utilization: New estimates from 40 countries. Emergency Medical Journal, 33:794–800, 2016.

[10] Pieter-Tjerk de Boer, Dirk P Kroese, Shie Mannor, and Reuven Y Rubinstein. A tutorial on the cross-entropy method. Annals of operations research, 134(1):19–67, 2005. ISSN 0254-5330.

[11] E. Erkut, A. Ingolfsson, and G. Erdogan. Ambulance location for maximum survival. Naval Research Logistics, 55(1):42–58, 2008. doi: 10.1002/nav.20267. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-38849174747&doi=10.1002%2fnav.20267&partnerID=40&md5=52db3ffe9fb55ca2cd44a46c40bef586.

[12] Reza Zanjirani Farahani, Samira Fallah, Rubén Ruiz, Sara Hosseini, and Nasrin Asgari. Or models in urban service facility location: A critical review of applications and future developments. European journal of operational research, 276(1):1–27, 2019.

[13] Andrew Fraser, Jessica Newberry Le Vay, Peter Byass, et al. Time-critical conditions: assessment of burden and access to care using verbal autopsy in agincourt, south africa. BMJ Global Health, 5:e002289, 2020.

[14] Johan Holmén, Johan Herlitz, Sven-Erik Ricksten, Anneli Strömsöe, Eva Hagberg, Christer Axelsson, and Araz Rawshani. Shortening ambulance response time increases survival in out-of-hospital cardiac arrest. Journal of the American Heart Association, 9(21):e017048, 2020.

[15] Craig Hooper, Jamie Ranse, and Alison Hutton. How is ambulance patient care and response time data collected and reported in malaysia and indonesia? Australasian Journal of Paramedicine, 16:1–8, 2019. doi: 10.33151/ajp.16.683. URL https://doi.org/10.33151/ajp.16.683.

[16] Badan Pusat Statistik Provinsi DKI Jakarta. Migrasi Penduduk JABOTABEK 2001. Badan Pusat Statistik provinsi DKI Jakarta, 2001.

[17] Badan Pusat Statistik Provinsi DKI Jakarta. DKI Jakarta Province in Figures. Badan Pusat Statistik provinsi DKI Jakarta, 2020.

[18] Badan Pusat Statistik Provinsi DKI Jakarta. Analisis Profil Penduduk Provinsi DKI Jakarta: mendeskripsikan Peran Penduduk dalam Pembangunan. Badan Pusat Statistik provinsi DKI Jakarta, 2023.

[19] V.A. Knight, P.R. Harper, and L. Smith. Ambulance allocation for maximal survival with heterogeneous outcome measures.

Omega, 40(6):918–926, 2012. doi: 10.1016/j.omega.2012.02.003. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-84863413219&doi=10.1016%2fj.omega.2012.02.003&partnerID=40&md5=c4ba0b07a8b1894bde6a237cc2242a4f.

[20] Xueping Li, Zhaoxia Zhao, Xiaoyan Zhu, and Tami Wyatt. Covering models and optimization techniques for emergency response facility location and planning: a review. Mathematical Methods of Operations Research, 74:281–310, 2011.

[21] Yang Liu, Yun Yuan, Jieyi Shen, and Wei Gao. Emergency response facility location in transportation networks: A literature review. Journal of traffic and transportation engineering (English edition), 8(2):153–169, 2021.

[22] Judy A Lowthian, Peter A Cameron, Johannes U Stoelwinder, Andrea Curtis, Alex Currell, Matthew W Cooke, and John J McNeil. Increasing utilisation of emergency ambulances. Australian Health Review, 35(1): 63–69, 2011.

[23] Richard McCormack and Graham Coates. 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, 2015.

[24] Laura A McLay and Maria E Mayorga. Evaluating emergency medical service performance measures. Health care management science, 13: 124–136, 2010.

[25] Jennifer Sian Morgan, Susan Howick, and Valerie Belton. A toolkit of designs for mixing discrete event simulation and system dynamics. European Journal of Operational Research, 257(3):907–918, 2017.

[26] Geraint I Palmer, Vincent A Knight, Paul R Harper, and Asyl L Hawa. Ciw: An open-source discrete event simulation library. Journal of Simulation, 13(1):68–82, 2019.

[27] Michael Pidd. Computer simulation in management science. Number 5th. John Wiley and Sons Ltd, 2004.

[28] Virginia Plummer, Malcolm Boyle, et al. Ems systems in lower-middle income countries: a literature review. Prehospital and disaster medicine, 32(1):64–70, 2017.

[29] Aryono D Pusponegoro. Terrorism in indonesia. Prehospital and disaster medicine, 18(2):100–105, 2003.

[30] Melanie Reuter-Oppermann, Pieter L van den Berg, and Julie L Vile. Logistics for emergency medical service systems. Health Systems, 6(3): 187–208, 2017.

[31] Reuven Rubinstein. The cross-entropy method for combinatorial and continuous optimization. Methodology and computing in applied probability, 1(2):127–, 1999. ISSN 1387-5841.

[32] İshak Şan, Eren Usul, Burak Bekgöz, and Semih Korkut. Effects of covid-19 pandemic on emergency medical services. International Journal of Clinical Practice, 75(5):e13885, 2021.

[33] The Ciw library developers. Ciwpython/ciw: v2.3.3. Dec 2022. doi: 10.5281/zenodo.7407301.

[34] T.D. Valenzuela, D.J. Roe, G. Nichol, L.L. Clark, D.W. Spaite, and R.G. Hardman. Outcomes of rapid defibrillation by security officers after cardiac arrest in casinos. New England Journal of Medicine, 343(17):1206–1209, 2000. doi: 10.1056/NEJM200010263431701. URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-0034718964&doi=10.1056%2fNEJM200010263431701&partnerID=40&md5=c10f7beb73c77cd6fec29db7d76cc995.

[35] Alexander Veser, Florian Sieber, Stefan Groß, and Stephan Prückner. The demographic impact on the demand for emergency medical services in the urban and rural regions of bavaria, 2012–2032. Journal of Public Health, 23:181–188, 2020.

[36] Wei Wang, Shining Wu, Shuaian Wang, Lu Zhen, and Xiaobo Qu. Emergency facility location problems in logistics: Status and perspectives. Transportation research part E: logistics and transportation review, 154: 102465, 2021.

[37] Liga Yusvirazi, Andi Ade Wijaya Ramlan, and Peter C Hou. State of emergency medicine in Indonesia. Emergency Medicine Australasia, 30 (6):820–826, 2018.