Ambulance Emergency Response Optimization in Developing Countries

Justin J. Boutilier, Timothy C.Y. Chan
Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Ontario M5S 3G8, Canada
j.boutilier@mail.utoronto.ca  tcychan@mie.utoronto.ca

The lack of emergency medical transportation is viewed as the main barrier to the access and availability of emergency medical care in low and middle-income countries (LMICs). In this paper, we present a robust optimization approach to optimize both the location and routing of emergency response vehicles, accounting for uncertainty in travel times and spatial demand characteristic of LMICs. We traveled to Dhaka, Bangladesh, the sixth largest and third most densely populated city in the world, to conduct field research resulting in the collection of two unique datasets that inform our approach. This data is leveraged to develop machine learning methodologies to estimate demand for emergency medical services in a LMIC setting and to predict the travel time between any two locations in the road network for different times of day and days of the week. We combine our robust optimization and machine learning frameworks with real data to provide an in-depth investigation into three policy-related questions. First, contrary to standard practice in high-income countries, we demonstrate that outpost locations optimized for weekday rush hour lead to good performance for all times of day and days of the week. Second, we find that significant improvements in emergency response times can be achieved by re-locating a small number of outposts and that the performance of the current system could be replicated using only one-third of the resources. Lastly, we show that a fleet of small motorcycle-based ambulances has the potential to significantly outperform traditional ambulance vans. In particular, they are able to capture three times more demand while reducing the median response time by 42% due to increased routing flexibility offered by more nimble vehicles on a larger road network. Our results provide practical insights for emergency response optimization that can be leveraged by hospital-based and private ambulance providers in Dhaka and other urban centers in developing countries.

Key words: Robust optimization, machine learning, facility location, global health, emergency medicine.

1. Introduction

Time-sensitive medical emergencies are a major health concern in low and middle income countries (LMICs), comprising one third of all deaths ([Razzak and Kellerman 2002]). Examples of such emergencies include cardiac arrest, motor vehicle accidents, and maternal health issues such as childbirth. Over the last decade, researchers and international organizations have stressed the need for increased focus on emergency medical care in LMICs. ([United Nations 2010], [World Health])
In particular, the 66th World Health Assembly passed a resolution (60.22) that “recognizes the necessity of evidence-based approaches to development of emergency care and asks WHO to promote emergency medicine research” (Anderson et al. 2012). However, despite widespread evidence that emergency medical care in LMICs save lives (Sodemann et al. 1997, Schmid et al. 2001), poor access and availability continues to be a major problem (Kobusingye et al. 2005, Levine et al. 2007) with the lack of emergency medical transportation noted as being the main barrier (Lungu et al. 2001, Macintyre and Hotchkiss 1999).

Optimizing the transport of emergency patients in urban centers in LMICs comes with unique challenges that are not present in high-income countries. First and foremost, traffic can be extremely unpredictable, and route disruptions caused by political demonstrations or extreme congestion occur regularly (Jain et al. 2012, Pojani and Stead 2015). Second, it is not the norm, and often not possible due to congestion, for motorists to yield for emergency vehicles. As a result, route optimization (and vehicle outpost location, by extension) becomes a critical component for improving emergency vehicle response times. Third, LMICs generally do not have historical emergency call data that can be used to forecast future emergency demand. In fact, most LMICs do not have a centralized emergency response system, so the prospect of collecting a large, high-quality dataset is itself a major challenge. Together, these challenges lead to a high degree of uncertainty in both travel times and spatial demand. The nature of these uncertainties directly impacts any modeling approach, which must be compatible with “small data” environments characteristic of LMICs.

In this paper, we develop a robust optimization approach to optimize both the location and routing of emergency response vehicles, accounting for uncertainty in travel times and spatial demand characteristic of LMICs. We traveled to Dhaka, Bangladesh, the sixth largest and third most densely populated city in the world, to conduct field research resulting in the collection of two unique datasets that inform our approach. First, we obtained a field dataset that includes patient travel data associated with several thousand hospital arrivals. This data, acting as a proxy for historical call data available in all modern, high-income countries, is leveraged to develop a machine learning methodology for estimating demand for emergency medical services in a LMIC setting. Second, we equipped five vehicles with custom-built GPS devices that recorded their time and location over a period of 30 days, allowing us to understand traffic and road network characteristics in Dhaka. We then developed a machine learning framework that uses the GPS data, along with census data, to predict the travel time between any two locations in the road network for different times of day and days of the week. For both demand and travel times, our predictions are leveraged to create data-driven uncertainty sets that are input into our robust location-routing model.

Like many urban centers in developing countries, Dhaka does not have a fleet of ambulances that form a centralized emergency response system. Instead, patients use a variety of transportation modes to reach hospitals in emergencies, including rickshaws, compressed natural gas (CNG)
vehicles (i.e., three-wheeled motorcycles), private cars, and private or hospital-based ambulance services. Our modeling framework is well-suited to handle different transportation modes, which are accounted for via differences in road network connectivity according to vehicle type. Smaller and more nimble vehicles can traverse roads that larger vehicles cannot access. Therefore, the consideration of transportation mode affects the ultimate computational tractability of our models. In this paper, we focus on traditional ambulance vans and smaller response vehicles based on the locally inspired CNGs, which have platforms that can be used for patient transport. Three-wheeled ambulances have been recently proposed in Bangladesh [Wadud 2017], but are not yet implemented and their potential impact on response times has not been studied in the scientific literature.

Ambulance services in Dhaka are currently decentralized, meaning there are both private ambulance service providers, which are for-profit businesses, and ambulance fleets that belong to hospitals. Both types of organizations are incentivized to increase the number of patient transports they make, but lack appropriate decision support tools to optimize their operations. For example, hospitals do not currently strategically pre-position their ambulances in the city, but rather position their entire fleet at the hospital. Therefore, private or hospital-based ambulance services are natural knowledge users of our research. Until recently, contact information for these services was also decentralized and unique to each provider, providing significant access challenges for patients. However, in December 2017, Bangladesh introduced the first centralized emergency services number “999” [Dhaka Tribune 2017]. The insights derived from our results can inform government policy on how to build a centralized emergency response system and aid non-government organizations to determine how to best position emergency response vehicle outposts. In particular, we use our real data and our modeling framework to answer four policy-related questions and derive practical insights for emergency response optimization in Dhaka and other LMICs:

1. **Should different outpost locations be used for different times of day?** (Section 6.1)
   While it is common practice in high-income countries for ambulance locations to be adjusted spatiotemporally throughout the day, does that value persist in LMICs?

2. **What performance improvements are possible by optimizing outpost locations?** (Section 6.2)
   How different would a centralized, optimized system be from the current situation where ambulances are parked at hospitals? How does re-positioning outpost locations compare to adding new locations?

3. **How much can the system be improved by using CNG-based response vehicles?** (Section 6.3)
   Can CNG-based ambulances capture additional demand that is currently unserved (or under-served) by existing ambulance vans? What is the potential value of increased routing flexibility offered by CNGs given their ability to traverse smaller roads in the network that are inaccessible to vans?
4. How important is it to consider uncertainty when designing an emergency response network? (Section 6.4) What is the performance improvement of our robust optimization model compared to a deterministic model? How does our robust approach compare to a perfect information approach?

The problem of optimizing emergency vehicle response has historically been cast as a facility location problem (Toregas et al. 1971). Although the facility location literature is rich, there is no unified framework for optimizing emergency vehicle response under both edge-based travel time uncertainty and demand uncertainty (Ahmadi-Javid et al. 2017). A key distinction between this paper and previous work is how we model travel time uncertainty. Our model provides a general edge-based framework for travel time uncertainty, whereas previous research has focused on modeling travel time uncertainty using a path-based approach (Snyder 2006). Edge-length uncertainty is critical for our model because many of the underlying causes of travel time uncertainty in Dhaka (e.g., intersections without signal control, floods, strikes, etc.) impact small subsets of edges as opposed to entire routes in isolation. Uncertainty on individual edges can affect multiple routes and must be accounted for during optimization. Our routing problem is effectively a robust shortest path problem and, depending on how we model edge-length uncertainty, is equivalent to a regularized shortest path problem. The equivalence between robustness and regularization has been noted in domains such as regression (Bertsimas and Copenhaver 2017), but has not been previously demonstrated for the shortest path problem.

Overall, we use the aforementioned challenges faced by LMICs and gaps in the facility location literature to motivate the development of a novel location-routing model that is tailored for emergency response optimization in developing urban centers. We make the following contributions:

- We develop a methodology to predict emergency demand spatially for urban centers without historical demand data by decomposing demand into components that can be estimated using census data and machine learning models. This approach represents the first attempt to predict emergency demand in a developing urban center. Our complete dataset, including census, survey, and hospital location data is unique because, to the best of our knowledge, hospital arrival surveys and patient travel data have never been collected together previously in any LMIC (Section 3).

- We develop and compare several machine learning models to predict travel time on the Dhaka road network by time of day and day of week, using a dataset of vehicle trips collected by our custom-made GPS devices. We find that a random forest model performs the best, with a 43.3–64.2% improvement over several baseline approaches. This paper is the first to use real travel time data from a LMIC for optimization (Section 4).

- We develop a novel edge-based reformulation of the classical path-based $p$-median problem. This reformulation forms the foundation of a two-stage robust optimization model that considers
both uncertain edge lengths (travel time) and node weights (demand). Our approach generalizes previous emergency facility location models based on the $p$-median architecture and provides a unified framework for emergency response optimization under travel time and demand uncertainty that is suitable for LMICs. We develop several approaches to solve our model. First, we develop an equivalent a single-stage mixed-integer linear optimization problem. Second, we develop an exact scenario (i.e., row and column) generation algorithm that can improve the solution time by several orders of magnitude. For application to large-scale problems representative of the real road network in Dhaka, we develop a novel heuristic algorithm by extending a state-of-the-art $p$-median heuristic to work with edge-length uncertainty (Section 5).

- Using our framework and real data from Dhaka, we provide an in-depth investigation into the four policy-related questions posed above (Section 6):

1. In contrast to developing countries where there are significant gains from re-positioning ambulances according to the time and day, there is little to gain in Dhaka by optimizing outpost locations spatiotemporally. Instead, using outpost locations optimized for weekday rush hour leads to good performance for all times of day and days of the week.

2. A significant improvement in emergency response times can be achieved by re-locating a small number of outposts. For example, re-locating 15% of the outposts, a 34.6% reduction in average response time can be achieved. Furthermore, if a centralized network was designed from a clean slate, the performance of the current system could be replicated using roughly one-third of the resources currently in use.

3. A fleet of CNG-based ambulances has the potential to significantly outperform traditional ambulance vans. In particular, they can capture over three times the demand as ambulances while reducing the median response time by 42% due to increased routing flexibility. This gain requires emergency response providers to tailor outpost locations specifically for CNG-based ambulances, instead of locating them at outposts optimized for traditional ambulance vans.

4. Our robust solutions can reduce the median and worst-case response times by up to 33.0% and 45.8%, respectively, compared to a deterministic solution that does not take uncertainty into account. Furthermore, the performance of the robust solution is comparable to a solution that has access to perfect information on the uncertainty.

2. Literature review

Our work is related to three major streams of literature: 1) demand prediction in the context of emergency response optimization, 2) vehicle travel time prediction, and 3) facility location.
2.1. Demand prediction

While most papers use historical emergency call data as a direct estimate for future demand, a growing and more relevant body of literature uses that data to develop machine learning models that can predict future demand. Early approaches considered only spatial demand, using multiple linear regression to relate the magnitude of demand for ambulances with population and other socio-economic factors (e.g., Schuman et al. 1977, Kamenetzky et al. 1982). Key covariates can be summarized into three main groups: measures of population (e.g., household size), measures of economic status (e.g., employment rate, poverty level), and measures of social status (e.g., literacy rate, marriage rate). Temporal-only approaches were developed to forecast emergency calls at various time scales, including daily (Baker and Fitzpatrick 1986), multi-hour blocks (Trudeau et al. 1989), and hourly (Matteson et al. 2011). Finally, there exist methods to predict future emergency demand at fine spatiotemporal resolutions (Setzler et al. 2009, Zhou et al. 2015).

The aforementioned approaches rely on granular historical call data to train prediction models. High-income countries tend to be data-rich, so research efforts have focused on advanced demand prediction techniques using this abundant and granular data. However, in most LMICs, historical call data is not available (Bradley et al. 2017). In this paper, we develop a new approach that does not use historical call data and instead makes use of the limited spatiotemporal data available in many LMICs.

2.2. Travel time prediction

Initial research on predicting edge-based travel times for ambulances focused on developing non-linear relationships between travel time and distance, but did not account for the time of day, day of week, or spatial location of trip (Kolesar et al. 1975, Budge et al. 2010). Recent Bayesian approaches address these limitations (Hofleitner et al. 2012a,b, Westgate et al. 2016). However, all prior research depends directly on the availability of historical emergency transport data, which typically does not exist in LMICs.

In recent years, machine learning approaches have gained popularity and demonstrated superior prediction accuracy for general travel time estimation (Vlahogianni et al. 2014). Travel times for emergency vehicles and regular vehicles are similar in LMICs because road users do not yield for ambulances. As a consequence, we employ a general travel time prediction approach similar to that of Zhang and Li (2015), who use a random forest model that accounts for distance, time of day, and day of week. We extend their model by incorporating demographic and geographic characteristics for the origin and destination nodes, which encodes spatial information about the trip.
2.3. Facility location

Facility location is a very well-studied field and we provide only a brief review of the relevant literature. For a general review of facility location, please see Owen and Daskin (1998) or Melo et al. (2009), and for a comprehensive review of facility location in the context of emergency medical services, please see Li et al. (2011), Basar et al. (2012), or Ahmadi-Javid et al. (2017).

2.3.1. Emergency response. Facility location models have been applied extensively to emergency medical services location problems with the majority of previous research focusing on ambulances. There have been many papers that investigate ambulance response optimization in urban areas in high-income countries (e.g., Brandeau and Larson 1986, Ingolfsson et al. 2008), in rural areas in high-income countries (e.g., Adenso-Diaz and Rodriguez 1997, Chanta et al. 2014), and in rural areas in LMICs (e.g., Bennett et al. 1982, Eaton et al. 1986). However, there have been only a few papers that consider urban areas in LMICs (Fujiwara et al. 1987, Basar et al. 2011, Salman and Yucel 2015, Zhang and Li 2015), and they differ from our work in several important aspects. First, these papers focus on upper-middle-income countries (China, Thailand, and Turkey), whereas we focus on a low-income country (Bangladesh). Second, previous urban ambulance response optimization research, including the papers listed above, has focused exclusively on regions that already have a centralized ambulance system. In contrast, our paper is the first to focus on a developing urban center without an existing ambulance system, which leads to new policy questions not considered in areas with an existing system.

2.3.2. Demand and travel time uncertainty. Demand uncertainty has received significant attention in general location-allocation problems (e.g., Shen et al. 2003, Atamturk and Zhang 2007, Baron et al. 2011) as well as in the specific context of ambulance response optimization (Beraldi et al. 2004, Beraldi and Bruni 2009, Noyan 2010). The ambulance-specific papers all use chance constraints to model uncertain demand, whereas we employ a scenario-based approach that integrates a machine learning model trained with our field data.

Travel time uncertainty in the context of ambulance response optimization has been focused on path-length uncertainty (Ingolfsson et al. 2008, Berman et al. 2013, Abdual Ghani and Ahman 2017). For networks with edge-length uncertainty, previous research has focused on the 1-median problem (Carson and Batta 1990, Averbakh 2003) and networks with special structure (Mirchandani and Odoni 1979, Mirchandani and Oudjit 1980). We are the first to investigate edge-length uncertainty for the general p-median problem applied to ambulance response optimization.

Nearly all previous literature on combining both edge-length and node-weight uncertainty has focused on the special case of the 1-median problem (Chen and Lin 1998, Vairaktarakis and Kouvelis 1999), whereas we develop a methodology for the general p-median problem under uncertainty. The
study by Serra and Marianov (1998), which considers the $p$-median problem with both uncertain path lengths and node weights, is the closest to our work. In contrast, we consider uncertain edge lengths and node weights, which can be interpreted as a generalization of their model.

3. Demand for emergency transportation

In this section, we provide a descriptive analysis of our census and survey data (Section 3.1) and develop a methodology to forecast spatiotemporal demand without historical call data (Section 3.2).

3.1. Descriptive analysis

3.1.1. Census data. We obtained the 2011 Dhaka census from the Bangladesh Bureau of Statistics. The census includes detailed demographic information for each of Dhaka’s 92 official wards (census tracts). Dhaka occupies a very small area of roughly 300 km$^2$ with a population of 8.95 million, which is a slightly larger population than New York City in under 40% of the area. Figure 1 illustrates the variation in four demographic characteristics across Dhaka’s 92 wards.

3.1.2. Survey data. We obtained data from 2,808 surveys administered by physicians to patients arriving at emergency departments (EDs) in 16 major hospitals (9 private, 7 government) across Dhaka. The survey had 14 questions (see EC.1) and was administered over 30 days between July 7, 2014 and August 25, 2014. The survey data includes the chief complaint, date, time, and ward location of the emergency, mode and cost of transportation, and the time of arrival at the ED. Our survey data is unique because it includes patient travel data and as a result, we are able to provide insights on the current emergency medical system that have not been previously captured.
Figure 2 (a) Number of trips and cost for each mode

Figure 2(b) Mode of transportation according to severity

Figure 2 Histograms of trip characteristics for each mode of transportation.

Figure 2(a) displays the various modes of transportation taken by patients and the costs incurred for each mode. Traditional ambulance vans were one of the least used modes of transport, comprising only 7.3% of all trips. Of the survey respondents who answered the question “Why did you not use an ambulance”, 16% indicated that they tried but it was not available and 7% cited slow response times, both of which are issues that can be addressed using our approach. Another major impediment was cost. Ambulances were found to be the most expensive mode of transportation with trips typically costing more than 16 US dollars (USD). For context, the average annual income in Bangladesh is 1,260 USD (Bangladesh Bureau of Statistics 2010). In contrast, rickshaws and CNGs, the two cheapest modes of transportation, comprised 34% and 25% of all trips, respectively.

Although overall ambulance utilization in Dhaka is low, Figure 2(b) shows that it is the mode with the highest proportion of trips for life-threatening emergencies. More than two thirds of all ambulance trips are for life-threatening emergencies, compared to less than one third of rickshaw trips. In life-threatening emergencies, ambulances become the third most common mode of transportation. This data suggests that patients recognize the importance of ambulances and are willing to use them for life-threatening emergencies. These findings also reinforce the importance of considering multiple vehicles types in LMICs. EC.2 provides further descriptive analyses of the data.

3.2. Emergency transport demand estimation

In this section, we outline our framework for estimating emergency demand using the limited data at our disposal. We do not have data on the total number of emergency transports as we would in North America because Dhaka does not have a centralized emergency medical system. Instead, we propose a novel decomposition of a standard metric, the annual number of emergency trips via mode $M$, into three components that can be estimated separately: population, per capita ED visits, and the proportion of ED visits that arrived via mode $M$. 
Let $T_{w,i}^M$ be a random variable representing the annual number of ED trips by individual $i$ originating in ward $w$ using mode $M$. For each individual, we model this random variable as an independent and identically distributed Poisson random variable with mean $\lambda_{w}^{M} = \lambda_{w} \delta_{w}^{M}$, where $\lambda_{w}$ represents the average annual number of ED visits per person from ward $w$ and $\delta_{w}^{M}$ represents the proportion of ED visits from ward $w$ that arrived via mode $M$. The annual number of ED trips originating in ward $w$ using mode $M$ is then $T_{w}^M = T_{w,1}^M + \cdots + T_{w,n_{w}}^M$, where $n_{w}$ represents the number of people in ward $w$. Since the individual $T_{w,i}^M$ are Poisson, $T_{w}^M$ is Poisson as well. The expected annual number of ED trips originating in ward $w$ using mode $M$, denoted $d_{w}^{M}$, is

$$d_{w}^{M} = \mathbb{E}(T_{w}^M) = n_{w} \lambda_{w}^{M} = n_{w} \lambda_{w} \delta_{w}^{M},$$

(1) since the $T_{w,i}^M$ are iid. Equation (1) suggests an approach to estimating $d_{w}^{M}$ by estimating its constituent terms $n_{w}$, $\lambda_{w}$, and $\delta_{w}^{M}$.

As written, equation (1) captures spatial heterogeneity in demand, but not temporal heterogeneity. Dhaka has a major difference in the spatial distribution of the daytime and nighttime population, due to daily migration. The magnitude of the daily migration out of Dhaka is estimated to be over 700,000 as many people leave the city during the day to work in the surrounding industrial areas. Thus, we refine equation (1) using $n_{w}^{DT}$ and $n_{w}^{NT}$, which represent the daytime population and nighttime population, respectively:

$$d_{w}^{M,DT} = n_{w}^{DT} \lambda_{w} \delta_{w}^{M},$$

(2)$$d_{w}^{M,NT} = n_{w}^{NT} \lambda_{w} \delta_{w}^{M}.$$ (3)

In total, there are four quantities to estimate, $n_{w}^{DT}$, $n_{w}^{NT}$, $\lambda_{w}$, and $\delta_{w}^{M}$. We summarize our estimates in this section and refer the reader to EC.3 for details regarding our approach to estimating each parameter. First, we use historical census data and a previously published rate of population growth to estimate the total 2016 daytime and nighttime population in Dhaka to be $n_{w}^{DT} = 8.24$ million and $n_{w}^{NT} = 8.95$ million, respectively. Figures 3(a) and 3(b) illustrate the estimated geographical distribution of the daytime and nighttime populations, respectively.

Leveraging published studies from other South Asian cities, we estimate the expected annual number of ED visits per capita in ward $w$ ($\lambda_{w}$) to lie in the interval $0.23 - 0.46$. Given data limitations and the coarseness of previous studies, we cannot generate ward-specific rates and instead settle on a single estimate for the city.

Finally, we train a regularized (lasso) logistic regression model for predicting the proportion of ED visits from ward $w$ arriving via mode $M$ ($\delta_{w}^{M}$). We combine our survey and census data to train our model. Our final features (see Tables EC.4 and EC.5), consistent with prior literature,
include measures of population (e.g., average household size), measures of social status (e.g., female
marriage rate), and measures of economic status (e.g., access to electricity). Figures 3(c) and 3(d)
display the model-predicted values of ED visits arriving via ambulance ($\delta^A_w$) and CNG ($\delta^C_w$), respectively. We find that areas of higher socioeconomic status are more likely to use ambulances as
compared to CNGs. For example, the wards with the largest values of $\delta^A_w$ include areas with a high
density of foreigners, government officials, and an area with major government offices, hospitals,
and universities. In contrast, CNG use is highest in many of the city’s outer wards, which include
slums. Putting all the pieces together, Figures 4(a) and 4(b) display the expected annual number
of daytime ($d^{A,DT}_w$) and nighttime ($d^{A,NT}_w$) ambulance trips in each ward assuming that $\lambda_w = 0.46$
(i.e., busiest case), while Figures 4(c) and 4(d) display the expected annual number of daytime
($d^{C,DT}_w$) and nighttime ($d^{C,NT}_w$) CNG trips, respectively.
4. Travel time analysis

In this section, we introduce two different road networks in Dhaka (Section 4.1), describe the travel time data collection methodology (Section 4.2), and develop machine learning models to predict the baseline travel time between any two locations in both networks (Section 4.3).

4.1. Road networks

The first road network that we consider is the ambulance network. In consultation with a transportation engineer in Dhaka and using a detailed map of the entire city, we determined exactly which roads are feasible for ambulance travel (many roads are too narrow for an ambulance). The ambulance network has 530 nodes and 1,280 edges. The second road network we consider is the complete network. This network – a superset of the ambulance network – includes all roads ranging from large arterial roads to small alleyways that can only be traversed by small vehicles like rickshaws, motorcycles, and CNGs. The complete road network has 5,358 nodes and 16,538 edges. Figure 5 displays both networks overlaid on Dhaka’s 92 wards.

4.2. Travel time data

We gathered vehicle location data using custom GPS devices and an accompanying Android mobile application developed by our collaborators. These devices were used by five volunteer citizens over 16 days from March 14, 2014 to June 13, 2014 and over 14 days from February 28, 2015 to April 2, 2015. All drivers were instructed to drive normally, using typical routes and speed. A map matching algorithm was developed to map the GPS data to edges on the road network (Ahmed et al., 2015). We obtained data for 269 unique trips. A trip is defined as a path through the network from some origin node to some destination node. A destination node is defined as either one from which
there is no subsequent GPS activity within 20 minutes on an edge emanating from that node or one with the last recorded GPS activity before the device was turned off by the driver. Trips ranged from 1 to 15 edges, with an average trip length of 4.1 edges. Edges in the network were present in a trip between 0 to 30 times, with an average of 3.9 observations per edge and a total of 1,103 edge observations (see Figure EC.5). The median travel time of a trip was 592 s (min: 10 s, max: 5543 s) while the median travel time on an edge was 105 s (min: 5 s, max: 5062 s).

To predict the travel time between two nodes, one challenge we face is limited data. In particular, if we use trip data to train our models, we are limited to only 269 observations. On the other hand, if we use edge data, which is more plentiful and includes 1,103 observations, then we are unable to capture the delays caused at nodes between edges (i.e., intersections) or the impact of traveling through a ward because most edges lie wholly within one ward. To deal with this trade-off, we develop a modified bootstrapping method that simultaneously solves the limited data issue and the issues with using edge data. This bootstrapping method expands our dataset by partitioning each trip into all contiguous sub-trips. For example, a trip that begins at node 1, visits nodes 2, then 3, and terminates at node 4 (denoted 1−2−3−4) would result in six sub-trips: 1−2, 2−3, 3−4, 1−2−3, 2−3−4, and 1−2−3−4. This bootstrapping process results in a total of 4,086 sub-trips, a 15 times increase in the size of the training set. The new sub-trip data is not a direct replication of trip data because each sub-trip has unique features according to the origin/destination of that sub-trip. Figure 6 displays a histogram of speeds for trips, edges, and sub-trips. Note that the sub-trip data includes both trip and edge data. The average speed (standard deviation) for the trip, edge, and sub-trip data is 2.05 km/h (3.90), 3.30 km/h (5.31), and 2.45 km/h (3.64), respectively.

4.3. Travel time prediction

Using the trip, edge, and sub-trip data, we compare four machine learning approaches for predicting the travel time in seconds between any two nodes (not necessarily adjacent) in the network. We
use 73 features including the distance on the road network between the given nodes, the day of week, and time of day. In addition, for both the origin and destination wards, we include building-type information (e.g., the number of commercial or industrial buildings) and the 27 demographic features used to predict emergency demand (see EC.3.3). The target is a real number that denotes the travel time in seconds between the two nodes. Two naive approaches serve as a baseline. The first, Naive S, predicts a constant equal to the average travel time from the empirical data, and the second, Naive D, is a simple linear regression model fit to distance only. We compare all six models using repeated 10-fold cross validation. See EC.4.2 for details.

Figure 7 displays a boxplot of the root mean squared error (RMSE) distribution across 100 repetitions for each of the prediction models tested on sub-trip data. The median RMSE for the Naive D (Naive S) approach was 648s (797s) when trained on edge data, 647s (731s) when trained on trip data, and 645s (684s) when trained on sub-trip data. The random forest model performed the best with a median RMSE of 629s, 605s, and 348s corresponding to improvements of 3% (21%), 6% (17%), and 46% (49%) over the Naive D (Naive S) approach when trained on edge, trip, and sub-trip data, respectively. All improvements were found to be statistically significant at $\alpha = 0.01$ using the Wilcoxon signed-rank test. For completeness, Figures EC.6 and EC.7 display boxplots of the RMSE for predictions tested on edge and trip data, respectively. These results depict a similar finding: a random forest model trained with sub-trip data is the most accurate.

A random forest model comprising 1,000 decision trees was selected as the final model and trained using all 4,086 sub-trips. Each feature was available for inclusion to all 1,000 trees and relative feature importance was determined using the number of trees in the forest to which that feature contributes. Table 1 lists the features that had a relative importance greater than 0.01. Our results suggest that travel distance, hour of day, and day of week are the three most important
features. As expected, travel distance is the most dominant feature with a relative importance of 0.4128. The hour of the day, which can be used as a proxy for peak traffic times, is the only other feature with an importance over 0.1. Our findings are consistent with the results of previous traffic studies, which also found travel distance and the time of day to be the main factors (Zhang and Li 2015, Vlahogianni et al. 2014). As mentioned in Section 2.2, our approach extends previous work by incorporating demographic features for the origin and destination nodes. We found nine geographical census features with a relative importance of at least 0.01. These additional features contribute to an 8% reduction in RMSE relative to a random forest model that only has access to distance and time features (see EC.4.2).

5. Optimization approach

In this section, we develop a two-stage robust optimization model to determine emergency response vehicle outpost locations. The model divides up the city into regions that optimize average, city-wide demand-weighted response time, where each region is served by one outpost. The outpost locations / regions are determined based on how vehicles will be routed from the outpost to demand points (second stage), considering uncertainty in both the demand and travel times.

We begin by introducing a novel edge-based location model that we prove to be equivalent to the classical $p$-median model. The advantage of our model is that it can handle edge-length uncertainty. Next, we introduce our models of uncertainty for emergency demand and travel times, based on the prediction approaches developed in Section 3.2 and 4.3. Finally, we develop and compare several solution approaches.

5.1. Network flow formulation

Let the road network be represented as the graph $G = (\mathcal{N}, \mathcal{E})$. Let $|\mathcal{N}| = n$, $|\mathcal{E}| = m$, and $A \in \mathbb{R}^{n \times m}$ denote the node-arc incidence matrix. Let $c \in \mathbb{R}^m$ denote the vector of edge lengths (i.e., travel
times) and \( d \in \mathbb{R}^n \) denote the vector of node weights (i.e., demand in terms of average annual emergency transports required). Let \( \alpha \in \mathbb{R}^n \) denote the supply available at each potential facility (i.e., number of trips that can be made from each outpost per year), \( P \) denote the number of outposts to be located, and \( e \) denote the vector of all ones. The decision variable representing the vector of flows along each edge is denoted by \( f \in \mathbb{R}^m \) (i.e., how many trips occur on each edge annually). The outpost location variable is given by \( y \in \{0,1\}^n \) where 1 indicates an outpost is located at node \( i \in \mathcal{N} \). In vector form, our deterministic network flow formulation (NFF) is:

\[
\text{NFF:} \quad \text{minimize} \quad c'f \\
\text{subject to} \quad e'y = P, \\
\quad Af \leq \alpha I y - d, \\
\quad f \geq 0, \\
\quad y \in \{0,1\}^n.
\]

(4)

To ensure that (4) is feasible for any value of \( P \), we require the following assumption.

**Assumption 1.** \( \alpha_i \geq \sum_{i=1}^{n} d_i, \forall i \in \mathcal{N} \).

This assumption states that each outpost has enough capacity to service the entire system (i.e., all demand nodes). We do not consider queuing in our model because our primary focus is to determine where to strategically locate emergency response outposts, rather than determining the total number of emergency response vehicles. As a result, this is not a limiting assumption. The following Lemma follows immediately from this assumption (proof omitted).

**Lemma 1.** There exists an optimal solution to NFF such that each demand node is assigned to exactly one outpost.

This result generally holds true for uncapacitated facility location models such as the \( p \)-median. Finally, using Lemma [1] we can show the equivalence between NFF and the \( p \)-median problem.

**Theorem 1.** A solution is optimal for NFF if and only if it is optimal for the \( p \)-median problem.

**Proof.** See EC.5

The proof of Theorem [1] provides a constructive approach to obtain an optimal solution of NFF given an optimal solution of the \( p \)-median problem, and vice versa. Mathematically, this approach provides a polynomial-time many-one reduction between the NFF and the \( p \)-median problem in both directions [Post 1944, Karp 1972].
5.2. Robust optimization model

In this section, we present our two-stage robust optimization model, considering both the travel times $c$ and demands $d$ as uncertain with $C$ and $D$ representing the corresponding uncertainty sets, respectively. Our general two-stage robust network flow formulation is:

$$\text{R-NFF}: \min_{y} \max_{c \in C,d \in D} \min_{f} \quad c'f$$

subject to

$$e'y = P,$$

$$Af \leq \alpha Iy - d,$$

$$f \geq 0,$$

$$y \in \{0,1\}^{n}. \quad (5)$$

The two-stage nature of our formulation is well-suited to the problem of emergency outpost location and vehicle routing. In the first stage, R-NFF determines the optimal outpost locations considering both $c$ and $d$ as uncertain. Intuitively, determining these locations is a high-level strategic decision that must be made under uncertainty, before demand or traffic are realized. Then, given the realized demand and travel time conditions, the second stage determines the optimal routes from the outposts to reach each demand point (i.e., patient location). Routing is a secondary decision that is used to inform the first stage location decision because the suitability of an outpost location is influenced by the route options emanating from that outpost.

5.2.1. Demand uncertainty set ($D$). To model uncertainty in emergency transport demand, we use a scenario-based uncertainty set. We use this approach to preserve tractability while still capitalizing on the richness of our demand predictions. To generate the scenarios that form the uncertainty set, we employ a form of bootstrapping and simulate possible realizations of demand vectors using our framework from Section 3.2.

We assume that the population in each ward follows a triangle distribution on the interval between the daytime and nighttime population, with a peak at the midpoint. Similarly, we assume that $\lambda_w$ follows a triangle distribution on the interval $[0.23 - 0.46]$, where the peak occurs at 0.40 for conservatism. Lastly, we assume that $\delta^M_w$, which is the probability of an emergency patient in ward $w$ taking mode $M$, follows a truncated normal distribution with a mean equal to the predicted ward value (Figures 3(c) and 3(d)) and a standard deviation equal to the median error (see Figures EC.3 and EC.4). For $N$ scenarios, the resulting uncertainty set is defined as $D = \{d^1, d^2, ..., d^N\}$, where the dimension of $d$ is equal to the number of nodes in the network.

The simulated demand vectors need to be adjusted so that the demand in each ward is spread proportionally to the road network nodes in that ward. In other words, we need to map the predicted demand based on the 92 wards to the $\sim 500$ or $\sim 5,000$ nodes in the ambulance and
complete networks, respectively. To make this adjustment, we generate a fine grid of nodes spaced 25m apart across all 92 wards, resulting in over 200,000 grid nodes. We distribute the simulated demand in each ward uniformly among the grid nodes in that ward. Then, we assign each grid node and its corresponding demand to the closest road network node using Euclidean distance.

5.2.2. Travel time uncertainty set (C). Uncertainty in travel time is modeled using an interdiction-based uncertainty set with an overall budget constraint [Wood 1993]. Intuitively, this set models an adversary who is adding traffic (i.e., increasing travel time) to the baseline traffic on each edge. The budget constraint restricts the total amount of travel time that can be added across the network. The mathematical formulation of this uncertainty set is \( C = \left\{ c_{ij}, (i,j) \in \mathcal{E} \mid c_{ij} = \hat{c}_{ij} + w_{ij}, \sum_{(i,j) \in \mathcal{E}} w_{ij} \leq B, w_{ij} \geq 0, \forall (i,j) \in \mathcal{E} \right\} \). We estimate the baseline travel time \( \hat{c}_{ij} \) for each edge using the final random forest model from Section 4.3. In our numerical experiments, we perform a detailed sensitivity analysis on the budget \( B \).

5.3. Solution Algorithms

In this section, we present several methods to solve R-NFF. First, we show that there is an equivalent single-stage mixed-integer optimization model for R-NFF. Then, we present an exact row-and-column generation algorithm to solve this equivalent problem. Finally, for the integer master problem, we devise an efficient heuristic that is needed for large-scale instances.

5.3.1. Equivalent mixed-integer optimization model. First, we replicate \( f \) for each of the scenarios in the demand uncertainty set \( \mathcal{D} \). Formally, we define \( f^k \) as the flow decision variable for scenario \( k = 1, \ldots, N \) and \( \lambda^k \) to be the dual variable corresponding to scenario \( k \) for the travel time uncertainty set constraint \( \sum_{(i,j) \in \mathcal{E}} w_{ij}^k \leq B \) in \( C \). The flow variable \( f^k \) corresponding to the limiting scenario for the first set of constraints in (6) is an optimal flow vector for (5).

**Theorem 2.** R-NFF is equivalent to the following mixed-integer linear optimization problem:

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad t \geq \hat{c}' f^k + \lambda^k B, \quad k = 1, \ldots, N, \\
& \quad A f^k \leq \alpha I y - d^k, \quad k = 1, \ldots, N, \\
& \quad f^k \leq \lambda^k e, \quad k = 1, \ldots, N, \\
& \quad f^k \geq 0, \quad k = 1, \ldots, N, \\
& \quad \lambda^k \geq 0, \quad k = 1, \ldots, N, \\
& \quad e'y = P, \\
& \quad y \in \{0,1\}^n.
\end{align*}
\]
**Proof.** See [EC.6]

Formulation (6) quickly becomes intractable as the number of scenarios increases and the size of the graph grows. We address these two challenges in the next two subsections. First, we develop a scenario generation algorithm that scales efficiently with the number of scenarios. Similar decomposition algorithms have been developed by Atamturk and Zhang (2007), Zeng and Zhao (2013), Gabrel et al. (2014) and Chan et al. (2017) for related two-stage problems. Second, we develop a heuristic to efficiently solve the master problem associated with the scenario generation approach.

### 5.3.2. Scenario Generation.

Consider a subset of the demand scenarios \( D_{|S|} = \{d_1, d_2, ..., d_{|S|}\} \subset D \), where \( S \) is an index set for the vectors in \( D_{|S|} \), and the corresponding relaxation of formulation (6) with \( D_{|S|} \) in place of \( D \):

\[
\text{R-NFF-MP:} \quad \begin{align*}
\text{minimize} & \quad y, t, f, \lambda \\
\text{subject to} & \quad t \geq \hat{c}^s f^s + \lambda^s B, \quad \forall s \in S, \\
& \quad A f^s \leq \alpha I \bar{y} - d^s, \quad \forall s \in S, \\
& \quad f^s \leq \lambda^s e, \quad \forall s \in S, \\
& \quad f^s \geq 0, \quad \forall s \in S, \\
& \quad \lambda^s \geq 0, \quad \forall s \in S, \\
& \quad e^t y = P, \\
& \quad y \in \{0,1\}^n.
\end{align*}
\]

(7)

The relaxed master problem, (7), produces a lower bound on the optimal value of (5) that can be tightened by adding additional scenarios to the set \( D_{|S|} \). Given an optimal solution \( \bar{y} \) to (7), we solve the following sub-problem, which is a linear optimization problem, for every \( d^k \in D \):

\[
\text{R-NFF-SP-k:} \quad Z^k_{SP} = \min_{f^k, \lambda^k} \hat{c}^k f^k + \lambda^k B \\
\text{subject to} & \quad A f^k \leq \alpha I \bar{y} - d^k, \\
& \quad f^k \leq \lambda^k e, \\
& \quad f^k \geq 0, \\
& \quad \lambda^k \geq 0.
\]

(8)

We choose the scenario \( k^* \in \arg \max_{k=1, \ldots, N} \{Z^k_{SP}\} \) and add the decision variables \( f^{k^*} \) and \( \lambda^{k^*} \), plus their corresponding constraints, to (7). Hence, this approach generates both rows and columns. The scenario generation algorithm terminates when the optimal value of (7) is equal to \( Z^*_{SP} \).

Finally, we comment on the structure of the subproblem (8) and connect it to a stream of research that draws an equivalence between robust optimization and regularization. Since \( \lambda^k \) is
being minimized in (8), the constraint $f^k \leq \lambda^k e$ identifies the maximum value of $f^k_{ij}$ over all $(i,j) \in \mathcal{E}$. Thus, we can rewrite (8) as (we drop the index $k$ for simplicity):

$$\begin{align*}
\text{minimize} & \quad \hat{c}' f + B \|f\|_\infty \\
\text{subject to} & \quad Af \leq \alpha \bar{y} - d, \\
& \quad f \geq 0.
\end{align*}$$

(9)

Formulation (9) is a “regularized” shortest path problem. Without the term $B\|f\|_\infty$ in the objective, (9) is exactly a shortest path problem. The extra term balances finding the shortest path with minimizing the maximum flow along any edge, which is weighted by the budget $B$. In our application, larger values of $B$ correspond to higher levels of traffic uncertainty. Thus, for large $B$, an optimal solution to (9) would prefer to spread out the flows (smaller maximum $f_{ij}$), forcing nature to expend more budget to “lengthen” multiple edges. Equivalently, if flows are concentrated on a few arcs, then nature has easy targets for adding traffic to cause maximal disruption. Our reformulation elucidates a clear connection between a robust shortest path problem and a regularized shortest path problem, similar to the way equivalences have been derived in regression (Xu et al. 2010, Bertsimas and Copenhaver 2017). For example, if we replace the constraint $\sum_{(i,j) \in \mathcal{E}} w_{ij} \leq B$ in $\mathcal{C}$ with $w_{ij} \leq B$, $\forall (i,j) \in \mathcal{E}$, then our subproblem is equivalent to a L1-regularized (lasso) problem.

5.3.3. Master problem heuristic. To solve the large-scale, real-world instances considered in our Dhaka experiments, we require a heuristic for the master problem, which is in essence a $p$-median problem. Although there are many heuristics that have been developed for the $p$-median problem, we cannot apply these algorithms directly because they are unable to handle edge-length uncertainty. Instead, we adapt the heuristic developed by Densham and Rushton (1992) for the classical $p$-median problem. This heuristic, designed for large-scale problems, leverages both the interchange heuristic proposed by Teitz and Bart (1968) and the alternate heuristic proposed by Maranzana (1964). A key benefit of this type of algorithm is that it scales well with both the size of the graph and the number of facilities ($P$). In fact, our heuristic represents the first tractable approach to solving large-scale instances of location problems with edge-length uncertainty. Our approach involves three main phases.

Initialization phase. We initialize our algorithm by randomly selecting $P$ nodes to serve as initial outpost locations, encoded by $\bar{y}$. We solve (8) with this $\bar{y}$ for every $d^k \in \mathcal{D}_S$, and identify $k^* \in \arg\max_{k \in S}\{Z^k_{SP}\}$, $f^{k^*}$, and $\lambda^{k^*}$. The corresponding cost of this solution is $\hat{c}' f^{k^*} + \lambda^{k^*} B$. An advantage of a random initialization phase is that our algorithm can be embedded in a meta-heuristic or a simple approach that considers multiple random starts. We investigate the impact of the number of random starts in our numerical experiments.
Interchange phase. In the interchange phase, we randomly swap a current outpost location node with a candidate node that is not currently in the solution. The new objective value is calculated as before after solving (8) for every $d^k \in D_S$. Swaps that reduce the objective value are accepted. We consider $\ell$ random interchanges, where $\ell$ is a user-chosen parameter.

Alternate phase. In the alternate phase, we use the incumbent solution from the interchange phase to partition the network into $P$ connected subgraphs that are disjoint from each other. Each subgraph contains exactly one outpost location and all demand nodes served by that outpost. We solve (7) for $P = 1$ (i.e., the robust 1-median problem) on each subgraph to determine the optimal outpost location. We then re-combine all subgraphs and the new optimal outpost locations to obtain an updated set of outpost locations, $\bar{y}$, in the full network. We compute the cost of this solution as before, by solving (8) for every $d^k \in D_S$. The alternate phase continues to partition and re-combine outpost locations until it has reached a local optimum. The algorithm then proceeds back to the interchange phase.

Termination. The algorithm iterates between the interchange and alternate phases until a solution from the alternate phase is found that does not result in any swaps during the interchange phase. The algorithm terminates with a solution to a single instance of the master problem (7).

Integration with scenario generation algorithm. The returned solution from the heuristic either terminates the scenario generation algorithm (when the the optimal value of (7) is equal to $Z_{SP}^k$) or is used as input to the sub-problem (8).

5.4. Comparison of solution approaches

In this section, we present results from a set of computational experiments that compare the effectiveness of our exact and heuristic scenario generation algorithms. To do so, we use smaller randomly-generated problem instances that can be solved to optimality.

5.4.1. Experimental setup. We use three random network instances to conduct our experiments. The first network has 30 nodes and 90 edges, the second has 50 nodes and 150 edges, and the third network has 75 nodes and 226 edges. For each graph, we vary the number of scenarios ($|S|$) in $D$, the interdiction budget ($B$) in $C$, and the number of vehicle outposts ($P$). Specifically, we consider: $|S| \in \{1, 10, 100, 1000, 10000\}$, $P \in \{1, 2, 5, 10, 25\}$, and $B \in \{0, 10, 50, 100, 250, 500, 1000\}$. In total, we solve 525 problem instances. We solve each instance using 1) a commercial solver (Gurobi), 2) our exact scenario generation algorithm (SGen), and 3) our heuristic scenario generation algorithm (HSGen) with 10 random starts and 10 interchanges. We chose this number of random starts and interchanges after testing our heuristic with different values (see EC.7). We set a maximum time limit of 36,000 seconds for each instance. All experiments were programmed using
MATLAB 2016a and run on a desktop computer with an Intel Core i7-4790K 4.0 GHz processor and 32 GB of RAM.

To estimate realistic edge-lengths for these instances, we randomly sample from the edge data distribution introduced in Section 4.2. To generate node-weights and demand scenarios, we used a modified version of the methodology outlined in Section 5.2.1. We estimate the population and $\lambda$ as in Section 5.2.1. Without underlying ward features, our logistic regression model is not applicable so we use the naive approach from EC.3.3 instead.

5.4.2. Scenario generation algorithm performance. The scenario generation algorithm was able to solve all 525 problem instances to optimality, while Gurobi struggled with larger instances. Table 2 scales the problem instances in terms of uncertainty set size. Gurobi was not able to solve any of the instances that had 100 or more demand scenarios, except the one with no travel time uncertainty. Table 3 compares the solution times for instances that vary in the size of the underlying network and the number of outposts located. The scenario generation approach enjoys the largest speed up for intermediate values of $P$.

5.4.3. Heuristic algorithm performance. Table 2 also compares the optimal cost and solution time between SGen and HSGen as a function of the number of scenarios and the interdiction budget. The objective function value is displayed as mean response time, in seconds. To determine this value, we divide the actual objective function value by the total number of trips. The performance of the heuristic algorithm remains relatively stable as the number of scenarios increases with solutions times that are an order or magnitude less than SGen. As the travel time budget increases, the gap between SGen and HSGen decreases, while the solution time of HSGen remains essentially constant. Table 3 compares the optimal cost and solution time between SGen and HSGen as a function of the size of the graph and the number of outposts, while holding both the interdiction budget and the number of scenarios constant at 100. Across all instances, the heuristic algorithm was able to obtain the optimal solution when the number of outposts was small. The performance also remains relatively stable as the size of the graph grows. However, the performance degrades as $P$ grows. This degradation in performance is balanced by up to an order of magnitude speed-up in certain cases. While HSGen does not close the optimality gap as the size of the problem increases, for the large-scale, real-world instances of the robust problems that we solve in Section 6, it is the only method capable of generating a solution in a reasonable time limit.

6. Application to Dhaka

In this section, we demonstrate the application of our models using data from Dhaka. Each of the following subsections addresses a policy question relevant to the design of an emergency response
Table 2  Comparison of objective function values and solution times between Gurobi, SGen, and HSGen as a function of graph size. The number of vehicles and the size of the graph are held constant at 5 and, 75 nodes and 226 edges, respectively.

| S  | Objective function value | Solution time |
|----|--------------------------|---------------|
|    | Budget | SGen | HSGen | Gap (%) | Gurobi (s) | SGen (s) | HSGen (s) |
| 1  | 1      | 133.4 | 142.8 | 6.6 | 3.4 | 3.4 | 3.7 |
|    | 10     | 134.8 | 141.0 | 4.4 | 4.3 | 4.3 | 3.6 |
|    | 100    | 145.9 | 171.7 | 15.0 | 5.6 | 5.7 | 3.7 |
|    | 1000   | 205.1 | 214.7 | 4.4 | 10.7 | 11.1 | 4.3 |
| 10 | 1      | 135.1 | 156.7 | 13.7 | 179.5 | 37.3 | 9.4 |
|    | 10     | 136.6 | 147.8 | 7.5 | 192.5 | 31.2 | 9.2 |
|    | 100    | 148.4 | 169.0 | 12.2 | 625.3 | 109.7 | 10.7 |
|    | 1000   | 208.3 | 209.8 | 0.7 | 4581.7 | 72.0 | 12.1 |
| 100| 1      | 143.2 | 164.9 | 13.1 | 26857.0 | 137.9 | 19.1 |
|    | 10     | 144.5 | 163.1 | 11.4 | - | 170.1 | 24.1 |
|    | 100    | 153.7 | 167.1 | 8.0 | - | 197.5 | 16.9 |
|    | 1000   | 218.4 | 227.2 | 3.9 | - | 2974.8 | 24.4 |
| 1000| 1     | 140.1 | 153.8 | 8.9 | - | 585.6 | 84.2 |
|     | 10     | 141.4 | 161.8 | 12.6 | - | 945.3 | 63.0 |
|     | 100    | 152.5 | 158.4 | 3.8 | - | 1160.0 | 93.7 |
|     | 1000   | 212.1 | 219.0 | 3.1 | - | 3721.8 | 34.8 |
| 10000| 1    | 142.5 | 162.4 | 12.2 | - | 3856.3 | 581.2 |
|      | 10     | 143.8 | 161.6 | 11.0 | - | 3264.3 | 462.7 |
|      | 100    | 154.7 | 164.0 | 5.7 | - | 9306.5 | 581.5 |
|      | 1000   | 217.8 | 227.0 | 4.0 | - | 5475.0 | 350.7 |

Table 3  Comparison of objective function values and solution times between SGen and HSGen as a function of graph size. The number of scenarios and the interdiction budget are held constant at 100 and 100, respectively.

| Nodes | Edges | P | Objective function value | Solution time |
|-------|-------|---|--------------------------|---------------|
|       |       |   | SGen | HSGen | Gap (%) | Gurobi (s) | SGen (s) | HSGen (s) |
| 1     | 1     | 213.9 | 213.9 | 0 | 87.8 | 12.7 | 2.0 |
| 2     | 155.8 | 155.8 | 0 | 82.2 | 25.5 | 9.2 |
| 5     | 93.0 | 102.8 | 9.6 | 473.9 | 25.1 | 7.7 |
| 10    | 46.1 | 60.3 | 23.5 | 210.4 | 94.8 | 42.9 |
| 25    | 7.3 | 11.3 | 35.4 | 10.6 | 142.1 | 40.1 |
| 1     | 287.0 | 287.0 | 0 | 525.1 | 38.5 | 9.6 |
| 2     | 191.0 | 191.0 | 0 | 1213.5 | 25.9 | 5.9 |
| 5     | 109.9 | 130.0 | 15.5 | 3534.0 | 89.6 | 25.4 |
| 10    | 62.8 | 88.7 | 29.2 | 3410.6 | 100.0 | 86.7 |
| 25    | 17.2 | 31.8 | 45.7 | 211.6 | 197.4 | 99.4 |
| 1     | 279.5 | 279.5 | 0 | 2671.8 | 31.7 | 6.9 |
| 2     | 232.2 | 232.2 | 0 | - | 88.8 | 13.1 |
| 5     | 153.7 | 167.1 | 8.0 | - | 197.5 | 16.9 |
| 10    | 99.4 | 137.3 | 27.6 | - | 400.9 | 32.0 |
| 25    | 40.4 | 65.9 | 38.7 | 2591.1 | 176.9 | 113.9 |
system: 1) Should different outposts be used for different times of day? (Section 6.1) 2) What performance improvements are possible by optimizing outpost locations? (Section 6.2) 3) How much can the system be improved by using CNG-based response vehicles? (Section 6.3) and 4) How important is it to consider uncertainty when designing an emergency response network? (Section 6.4).

For all experiments, we solve formulation (6) using the HSGen algorithm with 10 random starts and 10 random interchanges. Unless otherwise indicated, we use the uncertainty sets described in Sections 5.2.1 and 5.2.2 with 100 ambulance demand scenarios and a travel time budget \((B)\) of 1000 seconds. Through a detailed sensitivity analysis on the travel time budget (see EC.8.1), we find that optimizing outpost locations using a budget of 1000 seconds generates solutions that perform comparably to solutions optimized for other budgets. Sections 6.1, 6.2, and 6.4 use the ambulance road network, while Section 6.3 uses both the ambulance and complete road networks. All experiments were programmed using MATLAB 2016a and linear programming sub-problems were solved using Gurobi 7.0. The HSGen algorithm was able to solve each large-scale problem instance in under an one hour, and most were solved within 10 minutes. These real-world instances are comparable with the largest problems solved in the facility location literature and papers that focus on problems this large exclusively use heuristic methods (Fischetti et al. 2017). Given that we use a heuristic, the reported response time improvements should be seen as conservative estimates.

6.1. Should different outposts be used for different times of day?

In this section, we quantify the benefit of using different outpost locations for different times of day and days of the week, which we refer to as temporal snapshots. In many developed countries, demand is estimated at a fine spatiotemporal resolution, allowing ambulances to be re-positioned and response to be optimized for different snapshots (Zhou et al. 2015). These re-positioning strategies are motivated by changes in spatial demand as well as gaps in coverage caused by busy vehicles. While daily population migration is still a relevant motivation in LMICs, real-time re-positioning strategies to fill coverage gaps are unrealistic without a centralized emergency response system. Instead, there is a second key motivation for intra-day changes in ambulance locations in LMICs, which is the impact of changing traffic patterns on travel times. For example, we observed first-hand on several occasions during our field work the dramatic increase in travel times in different parts of the city during the evening rush hour. While traffic is less of a concern in high-income countries, emergency vehicles typically face the same traffic conditions as regular road users in LMICs since other vehicles do not (or cannot) yield to ambulances. Thus, our experiments in this section compare the performance of a system that changes outpost locations versus a configuration that keeps the ambulance outposts static.
Figure 8  The response time performance of outpost locations optimized for one specific snapshot and applied to other snapshots.

We use baseline travel times for three different temporal snapshots: weekday rush hour (Monday at 6pm), weekday overnight (Monday at 2am), and weekend midday (Saturday at 12pm). We use daytime population scenarios for rush hour and we use nighttime population scenarios for weekday overnight and weekend midday. For all three snapshots, we solve (6) with $P = 20$ and we test each set of outpost locations on the other two snapshots.

Figure 8 displays the distribution of average response time corresponding to outpost locations optimized for each of the three temporal snapshots. We find that ambulance response times are 30 minutes longer during rush hour as compared to both overnight and weekend. During rush hour, the rush hour-optimized locations have a median response time that is 5.0 min (9.5%) and 5.9 min (11.0%) faster than the average response time of the overnight- and weekend-optimized locations, respectively. During overnight and weekend, the rush hour-optimized locations are only 0.2 min (1.5%) and 1.1 min (6.1%) worse than the best outpost locations, respectively. Furthermore, when considering worst-case performance during rush hour, the rush hour-optimized locations are able to reduce response times by 5.6 min (10.6%) and 7.5 min (12.0%) over the overnight- and weekend-optimized locations, respectively.

6.1.1. Discussion and policy implications. Our results suggest that ambulance providers in Dhaka do not need to optimize outpost locations by time of day or day of week. Instead, providers can use static outpost locations optimized for daytime rush hour. The rush hour-optimized locations produce significant gains in response time during rush hour, while maintaining similar performance to specialized outpost locations at other times of the day. This finding is important because it supports reduced system complexity by removing the need to re-position emergency vehicles. In LMICs, it has been shown that complex solutions are far less likely to succeed compared to simple ones (Bradley et al. 2017). Thus, our rush hour-optimized solution is recommended since it is
optimal for the busiest time of day, close to optimal otherwise, and more likely to be implemented than a solution that involves regular re-positioning.

6.2. What performance improvements are possible by optimizing outpost locations?

Given the results in Section 6.1, we turn our attention to designing a static ambulance emergency response network for daytime rush hour and quantifying the gain from shifting away from the current practice of having hospital-based ambulances. There are currently 87 hospitals with emergency departments. Many of these hospitals have their own ambulance services, while others rely on private services. In both cases, ambulance providers typically position their fleets at the hospitals. To calculate response times, we assign each hospital to the closest node on the ambulance road network, resulting in 67 unique locations. Using these locations, Section 6.2.1 determines the baseline performance of the current hospital-based outpost locations in Dhaka.

We then consider three policy experiments for improving baseline ambulance response times that may inform the decision making of existing ambulance providers interested in improving or expanding their operations, as well as possibly new entrants or the government looking to design a system from scratch. In particular, Section 6.2.2 quantifies the value of re-positioning current outposts. Section 6.2.3 quantifies the value of adding additional outpost locations to the current network. Finally, Section 6.2.4 quantifies the performance of an ambulance network that is designed from scratch, without consideration of current outpost locations. The following sections measure performance of the ambulance networks on the rush hour snapshot, while EC.8.2, EC.8.3, and EC.8.4 confirm that similar results and insights hold for the other temporal snapshots.

6.2.1. What is the baseline performance of current hospital-based outpost locations? The median average response time of hospital-based outpost locations is 43.6, 10.0, and 13.2 minutes, during rush hour, overnight, and weekend, respectively. The variability in average response time across all 100 scenarios is much larger during rush hour, with a 21.7 minute difference between the best and worst scenarios, compared to 3.8 and 5.5 minute differences between the best and worst scenarios during overnight and weekend, respectively.

6.2.2. What is the value of re-positioning current outpost locations? We use a modified version of HSGen for these experiments. For each random start, we randomly choose the required number of outposts to re-position from the current locations and fix them for the remainder of the algorithm. As a result, the problem is reduced to determining the location of a specified number of outposts given a set of incumbent outpost locations.

Figure 9(a) displays the distribution of average response time for each number of re-positioned outposts. Re-positioning a small number of outposts generates a significant improvement in
response times, but we observe diminishing returns as the number of re-located outposts increases. For example, re-locating a single outpost reduces median response time by 8.0 minutes (18.4%) whereas re-positioning all 67 outposts provides an overall response time improvement of 22.0 minutes (50.5%).

Figure 10(a) shows a representative solution from the one-outpost repositioning problem. The current outpost locations are blue, the old outpost location is pink, and the re-positioned outpost location is yellow. Although the Euclidean distance between the old and new outpost locations is only 3.3 km, the time to travel between them is 403.8 minutes during rush hour, meaning that the new location can provide quicker service to an area that would otherwise see significant delays during rush hour. EC.8.2 presents similar results for weekend and overnight, suggesting that the southwest region is under-served for all three temporal snapshots.

6.2.3. What is the value of adding additional outpost locations to the current network? Figure 9(b) displays the distribution of average response time during rush hour as a func-
tion of outposts added. The addition of one new outpost provides the most value with a response time improvement of 8.0 minutes (18.4%), which is nearly the same as re-positioning a single outpost, suggesting that some current outposts provide minimal value. Figure 10(b) displays the location of a single additional rush hour, overnight, and weekend outpost. The additional weekend outpost is the same as the re-positioned outpost shown in Figure 10(a). Although the additional rush hour location is quite different, it is located in an area with many business, government offices, and universities; during rush hour, this area is particularly busy with people commuting home.

6.2.4. What is the value of designing a new emergency response network? Figure 9(c) displays the distribution of average response time for rush hour for newly optimized networks. We observe significant improvements in response time for the first few new outposts and observe steady gains up to about 20 new outposts. The response time performance of 20 new outposts is only 6.9 minutes (15.7%) worse than the current 67 outposts, suggesting that similar performance can be achieved with only one-third of the current resources. If we optimize a new network with 67 outposts, which is the same as re-positioning all 67 outposts, we can further improve response times but with significant diminishing returns. Figure 10(c) displays the location of 20 new outposts in relation to the current outposts. The new outposts are more strategically spread out compared to the current hospital locations that are concentrated in central Dhaka. For example, new outposts are added to the southwest and east of the city, which include low-income areas that were previously under-served by hospital-based outposts.

6.2.5. Discussion and policy implications. Our first two experiments (Sections 6.2.2 and 6.2.3) measure gains from local changes to the current network. The results suggest that a policy focused on re-positioning current outposts provides more value than a policy that aims to add additional outposts. For example, re-positioning 10 outposts provides roughly the same improvement in response time (15.1 minutes) as adding 10 new outposts (13.6 minutes). Practically, this result suggests that some of the current outpost locations are contributing very little to the overall response time calculation (i.e., they are rarely the fastest responding outpost to any given demand point) and that ambulance providers are better off re-positioning their entire fleet as opposed to sending a portion of their fleet to a new outpost location. In addition to providing significant response time reductions, a strategy focusing on re-positioning outposts is more likely to be cost-effective, which is a very important consideration resource-limited settings like LMICs.

If we consider a move towards centralization and a complete redesign of the current system, our third experiment shows that we can achieve roughly the performance of the current system with one-third of the resources. The non-governmental organization behind the newly implemented 999-number or a formal government agency seeking to implement a centralized emergency response
system may consider a complete re-design. Examining the 20 optimized locations from this experiment we find that nine of them coincide with hospital locations, while the other 11 are located off-site. Another way to view these results is that over 40 of the current hospital-based ambulance outposts can be removed without much impact on city-wide response times, or put to better use by concentrating the ambulances at fewer, more strategically located outposts.

Our experiments recommend putting outposts in the southwest, southeast, and northeast wards, suggesting that these areas are generally under-served. The southwest seems particularly underserved since both re-positioned and newly added outposts are located there. Knowing the demographics of the city, this result is not particularly surprising: the southwest wards form part of Old Dhaka and encompasses very dense low-income areas (see Figure 1) that have poor access to emergency transportation.

Overall, the key takeaway is that the current ambulance network in Dhaka is a dominated solution: response times in Dhaka can be significantly reduced without adding new resources, or equivalently, many fewer resources can be employed to match the current level of performance. Our modeling framework can play a pivotal role in the process to help decision makers strategically position their current ambulance resources. Of course, complementary initiatives will be required to achieve these gains, such as better public education about emergency medical transport and awareness of the newly created 999 number, which became operational in December 2017.

6.3. How much can the system be improved by using CNG-based response vehicles?

In this section, we consider the hypothetical situation where the city is served by a fleet of CNG-based emergency response vehicles that are able to traverse every road in the complete road network. Compared to the ambulance road network, the complete road network provides access to a larger portion of the city, including many dense low-income areas that are not accessible to ambulance vans. In some areas of the city, an entire sub-network of the complete road network is reduced to a single node in the ambulance road network. However, the distance between nodes in the sub-network and the nearest ambulance network node may be quite far, and it may be unrealistic to assume patients will coordinate multiple modes of transportation for different legs of their trip. As a result, we hypothesize that much of the emergency demand that arises from these low-income areas is lost or unserved. In Section 6.3.1, we quantify the potential emergency demand lost as a result of lack of access via the ambulance road network. To do this, we generate 100 demand scenarios using the prediction models for both ambulance and CNG demand from Section 3.2 and map this demand to nodes on the complete road network. Nodes that belong to the ambulance network retain the sum of the ambulance and CNG demand, while demand corresponding to complete network nodes that are not present in the ambulance network are assumed to be lost.
In addition to potentially capturing more demand, the complete road network also provides more routing options for smaller, CNG-based emergency response vehicles, which in turn may enable them to better avoid congestion and deal with travel time uncertainty. In Section 6.3.2, we quantify the value of increased routing flexibility provided by the complete network. We start by mapping demand (ambulance demand plus CNG demand) to nodes in the ambulance network. Then, we evaluate the response time performance of the current 67 hospital-based outpost locations as well as 20 new locations on both the ambulance and complete road networks. The 20 new locations are optimized for the corresponding road network, so they represent two distinct solutions.

**6.3.1. How much potential demand is lost by ambulance vans restricted the ambulance road network?** The complete network captures an average of 769,790 ambulance and CNG emergency trips per year, while the ambulance road network only captures 225,559 trips per year, representing a potential loss of 544,231 ambulance and CNG (70.7%) trips. Figure 11(a) visualizes the lost ambulance demand. The 530 green nodes are those that capture demand in both the ambulance and complete networks, while the 4,828 blue nodes only capture demand in the complete network and therefore, represent lost demand for the ambulance road network.

**6.3.2. What is the value of the increased routing flexibility offered by CNG-based response vehicles?** Figure 11(b) displays the response time performance for the current 67 baseline outpost locations and 20 new outpost locations on both the ambulance and complete road networks. The median average response time of the current 67 locations on the ambulance road network is 34.1 minutes with a range of 19.0 minutes between the best and worse scenarios. CNG-based ambulances located at the same outposts are able to reduce the median response time to 19.7 minutes (a 42.2% reduction) and reduce the min-max range to 10.5 minutes (a 44.9% reduction). If we consider 20 new outpost locations, we get a similar reduction in average response time of 41.5%, from 33.7 minutes to 19.7 minutes. We also see a similar decrease in variability across the 100 demand scenarios. Figure 11(c) shows the 20 new outpost locations for both the ambulance (pink nodes) and complete (yellow nodes) road networks. The four blue nodes represent outpost locations that are the same in both networks.

**6.3.3. Discussion and policy implications.** The results in this section represent the first attempt to provide quantitative evidence of the potential benefit of CNG-based ambulances in an LMIC. Recall from Section 3.1.2 that 23% of survey respondents indicated that ambulances were either not available or too slow to reach their location. CNG-based ambulances offer a potential solution to both these issues. Our results have three policy implications:
1. Smaller response vehicles can potentially capture three times more emergency demand than traditional ambulances vans in Dhaka. Much of the additional demand captured is generated from nodes in hard-to-reach and low-income areas, such as urban slums (southern and western clusters of nodes in Figure 11(a)). These areas are known to already suffer from poor access and availability of emergency medical care.

2. Smaller response vehicles are able to reduce the median average response time by 41.5-42.2%, while simultaneously reducing the worst-case by 42.5%. These reductions are entirely due to increased routing flexibility offered by having nimbler vehicles navigating a larger road network. These results may even be somewhat conservative because we did not incorporate the fact that CNG-based vehicles are typically able to travel faster than larger ambulances.

3. Our results demonstrate that the outpost locations chosen for CNG-based ambulances are very different from those chosen for traditional ambulances. This result emphasizes the importance of considering CNG-based ambulances independently; we cannot assume they should be positioned alongside traditional ambulances, even if the ambulance outpost locations are themselves optimized, because they are optimized for a different road network.

Overall, the key takeaway from these experiments is that CNG-based ambulances have the potential to not only significantly improve system efficiency through lower response times, but also simultaneously improve equity and access by capturing substantial demand in the hardest to reach areas of the city.

6.4. How important is it to consider uncertainty when designing an emergency response network?

In this section, we quantify the value of robustness by comparing our robust optimization model to the deterministic model (NFF), as well as to a perfect information formulation that solves NFF.
after the uncertainty has been realized. We focus on the situation where 20 new outposts are being located. We also examine how the performance gaps vary as the travel time budget is varied. The deterministic formulation uses the average demand and baseline travel times with no uncertainty, while the perfect information formulation finds a unique solution for each demand scenario.

Figure 12 displays the response time improvement of the robust and perfect information solutions over the deterministic solution for different levels of travel time uncertainty. For a travel time budget of 1000 seconds, the robust solution generates a 8.0% and 8.6% improvement over the deterministic solution in the median and worst-case average response time, respectively. Compared to the perfect information solution, the robust solution has a median average response time that is only 0.8% worse, with a worst-case average response time that is 1.9% better. As expected, the gains from the robust model increase as the size of the uncertainty set grows. For example, with a budget of 10,000 seconds, the robust solution improves upon the deterministic solution by 33.0% in median and 45.8% in worst-case average response time. At the same time, the performance of the robust solution continues to track the performance of the perfect information solution quite closely.

6.4.1. Discussion and policy implications. Our results demonstrate that a robust optimization framework tailored for the uncertainties faced by LMICs is able to produce solutions that significantly outperform solutions that do not consider uncertainty. As expected, the performance gains increase with the amount of uncertainty considered. Furthermore, our robust solutions are comparable to those derived from a perfect information model. Overall, these results further reinforce the importance of robustness for designing emergency response solutions in environments with substantial uncertainty characteristic of LMICs.
7. Conclusion

In this paper, we developed a comprehensive framework for emergency response optimization that combines two machine learning approaches with a robust optimization model tailored to address the specific challenges faced by LMICs. Our optimization model generalizes previous emergency response models in both high, middle, and low-income countries and provides a unified framework for emergency response optimization under travel time and demand uncertainty. We use two unique datasets that we collected in Dhaka, Bangladesh to train our machine learning models and build our uncertainty sets.

Using our real data and modeling framework, we address four policy questions related to the design of an emergency response system in LMICs, using Dhaka, Bangladesh as a target site. First, we demonstrated that population migration has a minimal impact on response times and that outpost locations optimized specifically for rush hour perform well throughout the day and week. Second, we demonstrated that re-positioning current outpost locations is a better alternative than adding new outposts, especially considering resource limitations in Dhaka. Additionally, a newly designed network can achieve comparable performance to the current network using only one-third of the resources. Half of the new outposts would coincide with current outpost locations, while the other half should be strategically positioned in the lower-income parts of the city. Third, we show that CNG-based ambulances may be able to capture three times more demand than ambulance vans due to their ability to access parts of the city with narrow roads such as slums. In addition, the routing flexibility offered by the larger road network available to CNGs can reduce average response times by 42% and worst-case response times by 43%. Our final experiment demonstrated that our robust optimization framework is able to produce networks with average response times that are up to 33.0% faster than a deterministic solution, comparable to a network designed with perfect information on the uncertainty.

Overall, this work highlights the opportunity of combining machine learning with optimization in the “small data” environments characteristic of LMICs to solve relevant healthcare and public policy challenges.

Acknowledgments

The authors gratefully acknowledge Dr. Moinul Hossain for his input on early stages of the project and for leading the data collection efforts. The authors thank Mehedi Hasan for his help with demand data collection, Mahfuzur Rahman Siddiquee for his help with travel time data collection, and all those who volunteered in Dhaka. We are grateful to Prof. Yu-Ling Cheng and Dr. Laurie Morrison for support and advice throughout this project. This research was supported by Grand Challenges Canada and the Natural Sciences and Engineering Research Council of Canada (NSERC).
References

Abdul Ghani, N., Ahmad, N. 2017. Analysis of MCLP, Q-MALP, and MQ-MALP with Travel Time Uncertainty Using Monte Carlo Simulation. Journal of Computational Engineering 2017.

Adenso-Diaz, B., Rodriguez, F. 1997. A simple search heuristic for the MCLP: Application to the location of ambulance bases in a rural region. Omega 25 181-187.

Ahmadi-Javid, A., Seyed, P., Syam, S.S. 2017. A survey of healthcare facility location. Computers and Operations Research 79 223-263.

Ahmed, N., Rahman Siddiquee, M., Karim, R., Zaman, M., Monzur, R., Hossain, M. 2015. Map Matching on Sparse GPS Data: A Perspective of a Developing City. 10th International Conference of Eastern Asia Society For Transportation Studies.

Anderson, P.D., Suter, R.E., Mulligan, R., Bodilwala, G., Razzak, J.A., Mock, C., International Federation for Emergency Medicine (IFEM) Task Force on Access and Availability of Emergency Care. 2012. World Health Assembly Resolution 60.22 and Its Importance as a Health Care Policy Tool for Improving Emergency Care Access and Availability Globally. Annals of Emergency Medicine. 60 35-44.

Atamturk, A., Zhang, M. 2007. Two-stage robust network flow and design under demand uncertainty. Operations Research 55 662-673.

Averbakh, I., 2003. Complexity of robust single facility location problems on networks with uncertain edge lengths. Discrete Applied Mathematics 127 505-522.

Baker, J.R., Fitzpatrick, K.E. 1986. Determination of an Optimal Forecast Model for Ambulance Demand Using Goal Programming. The Journal of the Operational Research 37 1047-1059.

Bangladesh Bureau of Statistics. 2010. Report of the Household Income and Expenditure Survey. Ministry of Planning.

Baron, O., Milner, J., Naseraldin, H. 2011. Facility location: A robust optimization approach. Production and Operations Management 20 772-785.

Basar, A., Catay, B., Unluyurt, T. 2011. A multi-period double coverage approach for locating the emergency medical service stations in Istanbul. Journal of the Operational Research Society 62 627-637.

Basar, A., Catay, B., Unluyurt, T. 2012. A taxonomy for emergency service station location problem. Optimization letters 6 1147-1160.

Bennett, V.L., Eaton, D.J., Church, R.L. 1982. Selecting sites for rural health workers. Social Science and Medicine 16 63-72.

Beraldi, P., Bruni, M.E., Conforti, D. 2004. Designing robust emergency medical service via stochastic programming. European Journal of Operational Research 158 183-193.

Beraldi, P., Bruni, M.E. 2009. A probabilistic model applied to emergency service vehicle location. European Journal of Operational Research 196 323-331.
Berchet, C. 2015. Emergency care services: trends, drivers, and interventions to manage the demand. Organization for Economic Co-operation and Development.

Berlin, G.N., ReVelle, C.S., Elzinga, D.J. 1976. Determining Ambulance and Hospital Locations for On-Scene and Hospital Services. Environment and Planning A 8 553-561.

Berman, O., Hajizadeh I., Krass, D. 2013. The maximum covering problem with travel time uncertainty. IIE Transactions 45 81-96.

Bertsimas, D., Copenhaver, M.S. 2017. Characterization of the equivalence of robustification and regularization in linear and matrix regression. European Journal of Operational Research. In Press.

Bradley, B.D., Jung, T., Tandon-Verma, A., Khoury, B., Chan, T.C.Y., Cheng, Y.L. 2017. Operations research in global health: a scoping review with a focus on the themes of health equity and impact. Health Research Policy and Systems 15 32.

Brandeau, M.L., Larson, R.C. 1986. Extending and applying the hypercube queueing model to deploy ambulances in Boston. National Emergency Training Center.

Budge, S., Ingolfsson, A., Zerom, D. 2010. Empirical analysis of ambulance travel times: the case of Calgary emergency medical services. Management Science 56 716-723.

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

Chan, T.C., Shen, Z.J.M., Siddiq, A. 2017. Robust facility location under demand location uncertainty. arXiv preprint arXiv:1507.04397.

Channouf, N., L’Ecuyer, P., Ingolfsson, A., Avramidis, A. 2007. The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta. Health Care Management Science 10 25-45.

Chanta, S., Mayorga, M.E., McLay, L.A. 2014. Improving emergency service in rural areas: a bi-objective covering location model for EMS systems. Annals of Operations Research 221 133-159.

Chen, B., Lin, C.S. 1998. Minmax regret robust 1-median location on a tree. Networks 31 93-103.

Densham, P.J., Rushton, G. 1992. A more efficient heuristic for solving large p-median problems. Papers in Regional Science 71 307-329.

Dhaka Tribune. 2017. ‘999’ emergency services begin. URL https://www.dhakatribune.com/bangladesh/2017/12/12/999-emergency-services-begin/.

Eaton, D.J., Hector, M.L. Sanchez, U., Morgan, J. 1986. Determining ambulance deployment in santo domingo, dominican republic. Journal of the Operational Research Society 113-126.

Fischetti, M., Ljubic, I., Sinnl, M. 2017. Redesigning Benders Decomposition for Large-Scale Facility Location. Management Science 63 2146-2162.
Fujiwara, O., Makjamroen, T., Gupta, K.K. 1987. Ambulance deployment analysis: a case study of Bangkok. European Journal of Operational Research 31 9-18.

Gabrel, V., Lacroix, M., Murat, C., Remli, N. 2014. Robust location transportation problems under uncertain demands. Discrete Applied Mathematics 164 100-111.

Goins, S., Conroy, M.B. 2015. New York State All Payer Emergency Room Visits. New York State Department of Health Statistical Brief.

Hakimi, S.L. 1964. Optimum locations of switching centers and the absolute centers and medians of a graph. Operations research 12 450-459.

Hakimi, S.L. 1965. Optimum distribution of switching centers in a communication network and some related graph theoretic problems. Operations Research 13 462-475.

Hamner, B., 2010. Predicting travel times with context-dependent random forests by modeling local and aggregate traffic flow. In: Data Mining Workshops (ICDMW), 2010 IEEE International Conference on IEEE 1357-1359.

Holliman, C.J., Mulligan, T.M., Suter, R.E., Cameron, P., Wallis, L., Anderson, P.D., Clem, K. 2011. The efficacy and value of emergency medicine: a supportive literature review. International Journal of Emergency Medicine 4 44.

Hofleitner, A., Herring, R., Bayen, A. 2012. Arterial travel time forecast with streaming data: A hybrid approach of flow modeling and machine learning. Transportation Research Part B 46 1097-1122.

Hofleitner, A., Herring, R., Abbeel, P., Bayen, A. 2012. Learning the dynamics of arterial traffic from probe data using a dynamic Bayesian network. IEEE Transactions on Intelligent Transportation Systems 13 1679-1693.

Hsia, R., Razzak, J.A., Tsai, A.C., Hirshon, J.M. 2010. Placing Emergency Care on the Global Agenda. Annals of Emergency Medicine 56 142-149.

Ingolfsson, A., Budge, S. and Erkut, E., 2008. Optimal ambulance location with random delays and travel times. Health Care management science 11 262-274.

Jain, V., Sharma, A., Subramanian, L. 2012. Road Traffic Congestion in the Developing World. Proceedings of the 2nd ACM Symposium on Computing for Development 11.

Jenelius, E., Koutsopoulos, H. 2013. Travel time estimation for urban road networks using low frequency probe vehicle data. Transportation Research Part B 53 64-81.

Kamenetzky, R.D., Shuman, L.J., Wolfe, H. 1982. Estimating Need and Demand for Prehospital Care. Operations Research 30 1148-1167.

Karim, M.Z., Hansen, E.L., Ahmad, B.U., Lahiry, S. 2009. A retrospective study of illness and admission pattern of emergency patients utilizing a corporate hospital in Dhaka, Bangladesh: 2006-2008. The ORION 32 1-7.
Karp, R.M. 1972. Reducibility Among Combinatorial Problems. Complexity of Computer Computations. Plenum, New York, 85-103.

Kobusingye, O.C., Hyder, A.A., Bishai, D., Hicks, E.R., Mock, C. and Joshipura, M. 2005. Emergency medical systems in low- and middle-income countries: recommendations for action. Bulletin of the World Health Organization 83 626-631.

Kolesar, P., Walker, W., Hausner, J. 1975. Determining the relation between fire engine travel times and travel distances in New York City. Operations Research 23 614-627.

Levine, A.C., Gadiraju, S., Goel, A., Johar, S., King, R., Arnold, K. 2007. International emergency medicine: a review of the literature. Acad Emerg Med. 14 182-1833.

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 74 281-310.

Lin, C.S., Chang, H., Shyu, K.G., Liu, C.Y., Lin, C.C., Hung, C.R., Chen, P.H. 1998. A method to reduce response times in prehospital care: the motorcycle experience. Am J Emerg. Med. 16 711-713.

Lungu, K., Kamfose, V., Hussein, J., Ashwood-Smith, H. 2001. Are bicycle ambulances and community transport plans effective in strengthening obstetric referral systems in Southern Malawi? Malawi Medical Journal 12 16-18.

Macintyre, K., Hotchkiss, D.R. 1999. Referral revisited: community financing schemes and emergency transport in rural Africa. Soc Sci Med. 49 1473-1487.

Mandell, M.B. 1998. Covering models for two-tiered emergency medical services systems. Location Science 6 355-368.

Maranzana, F.E. 1964. On the location of supply points to minimize transport costs. Operational Research Quarterly 15 261-270.

Matteson, D.S., McLean, M.W., Woodard, D.B., Henderson, S.G. 2011. Forecasting emergency medical service call arrival rates. Annals of Applied Statistics 5 1379-1406.

Melo, M.T., Nickel, S., Saldanha-Da-Gama, F. 2009. Facility location and supply chain management-A review. European journal of operational research 196 401-412.

Mirchandani, P.B., Odoni, A.R. 1979. Locations of medians on stochastic networks. Transportation Science 13 85-97.

Mirchandani, P.B., Oudjit, A. 1980. Localizing 2?medians on probabilistic and deterministic tree networks. Networks 10 329-350.

Neumann, J.v. 1928. Zur Theorie der Gesellschaftsspiele. Math. Annalen. 100 295-320.

Noyan, N. 2010. Alternate risk measures for emergency medical service system design. Annals of Operations Research 181 559-589.
Owen, S.H., Daskin, M.S. 1998. Strategic facility location: A review. European journal of operational research 111 423-447.

Pojani, D., Stead, D. 2015. Sustainable urban transport in the developing world: beyond megacities. Sustainability, 7 7784-7805.

PoSaw, L.L., Aggarwal, P., Bernstein, S.L. 1998. Emergency medicine in the New Delhi Area, India. Ann Emerg Med 32 609-615.

Post, E.L. 1944. Recursively enumerable sets of positive integers and their decision problems. Bulletin of the American Mathematical Society 50 284-316.

Raftery, K.A. 1996. Emergency medicine in southern Pakistan. Emerg Med 27 79-93.

Razzak, J.A., Kellerman, A.L. 2002. Emergency medical care in developing countries: is it worthwhile? Bulletin of the World Health Organization 80 900-905.

ReVelle, C.S., Swain, R.W. 1970. Central Facilities Location. Geographical Analysis 2 30-42.

Riley, P. 2000. Motorcycle medics. JEMS 25 32-41.

Royston, G. 2011. Meeting global health challenges through operational research and management science. Bulletin of the World Health Organization 89 683-688.

Royston, G. 2012. Special issue on global health. Health Care Manag Sci 15 171-174.

Salman, F.S., Ycel, E. 2015. Emergency facility location under random network damage: Insights from the Istanbul case. Computers and Operations Research 62 266-281.

Samai, O., Senegh, P. 1997. Facilitating emergency obstetrical care through transportation and communication, Bo, Sierra Leone. International Journal of Gynecology and Obstetrics 59 S157-164.

Schmid, T., Kanenda, O., Ahluwalia, I., Kouletio, M. 2001. Transportation for maternal emergencies in Tanzania: empowering communities through participatory problem solving. American Journal of Public Health 91 1589-1590.

Schuman, L.J., Wolfe, H., Sepulveda, J. 1977. Estimating demand for emergency transportation. Med Care 15 738-749.

Setzler, H., Saydam, C., Park, S. 2009. EMS call volume predictions: A comparative study. Computers and Operations Research 36 1843-1851.

Serra, D., Marianov, V. 1998. The p-median problem in a changing network: the case of Barcelona. Location Science 6 383-394.

Shen, Z.J.M., Coullard, C., Daskin, M.S. 2003. A joint location-inventory model. Transportation science 37 40-55.

Snyder, L.V. 2006. Facility location under uncertainty: a review. IIE Transactions 38 537-554.
Sodemann, M., Jakobsen, M.S., Molbak, K., Alvarenga, I.C., Aaby, P. 1997. High mortality despite good
care-seeking behaviour: a community study of childhood deaths in Guinea-Bissau. Bulletin of the World
Health Organization 75 205-212.

Streatfield, P.K., Karar, Z.A. 2008. Population Challenges for Bangladesh in the Coming Decades. J Health
Popul Nutr 26 261-272.

The Sixty-sixth World Health Assembly 2013. The World Health Organization

The World Health Report 2013: Research for universal health coverage. The World Health Organization.

Teitz, M.B, Bart, P. 1968. Heuristic methods for estimating generalized vertex median of a weighted graph.
Operations Research 16 955-961.

Toregas, C.R., Swain, R.W., ReVelle, C.S., Bergman, L. 1971. The location of emergency service facilities.
Operations Research 19 1363-1373.

Trudeau, P., Rousseau, J-M., Ferland, J.A., Choquette, J. 1989. An operations research approach for the
planning and operation of an ambulance service. Information Systems and Operational Research 27
95-113.

The Millennium Development Goals Report 2010. United Nations.

Vairaktarakis, G.L., Kouvelis, P. 1999. Incorporation dynamic aspects and uncertainty in 1-median location
problems. Naval Research Logistics 46 147-168.

Vile, J.L., Gillard, J.W., Harper, P.R., Knight, V. A. 2012. Predicting ambulance de- mand using singular
spectrum analysis. Journal of the Operations Research Society 63 1556-1565.

Vlahogianni, E.I., Karlaftis, M.G., Golas, J.C. 2014. Short-term traffic forecasting: Where we are and where
we are going. Transportation Research Part C 43 3-19.

Wadud, M. 2017. Cheap solar ambulances to speed into service in rural Bangladesh. Thomson
Reuters Foundation. URL: https://www.reuters.com/article/us-bangladesh-solar-ambulance/cheap-
solar-ambulances-to-speed-into-service-in-rural-bangladesh-idUSKBN15T1AP

Westgate, B.S., Woodard, D.B., Matteson, D.S., Henderson, S.G. 2016. Large-network travel time distribu-
tion estimation for ambulances. European Journal of Operational Research 252 322-333.

Wood, R.K. 1993. Deterministic network interdiction. Mathematical and Computer Modelling 17 1-18.

Xu, H., Caramanis, C., Mannor, S. 2010. Robust regression and Lasso, IEEE Transactions in Information
Theory 56 3561-574.

Zhang, Y., Chenyang, Y., Kan, H., Cao, J., Peng, Li., Xu, J., Wang, W. 2014. Effect of ambient temperature
on emergency department visits in Shanghai, China: a time series study. Environmental Health 13 1-8.

Zhang, Y., Haghani, A. 2015. A gradient boosting method to improve travel time prediction. Transportation
Research Part C 58 308-324.
Zhang, Z.H., Li, K. 2015. A novel probabilistic formulation for locating and sizing emergency medical service stations. Annals of Operations Research 229 813-835.

Zeng, B., Zhao, L. 2013. Solving two-stage robust optimization problems using a column-and-constraint generation method. Operations Research Letters 41 457-461.

Zhou, Z., Matteson, D.S., Woodard, D.B., Henderson, S.G., Micheas, A.C. 2015. A Spatio-Temporal Point Process Model for Ambulance Demand. Journal of the American Statistical Association 110 6-15.
Electronic Companion

EC.1. Hospital Survey

Table EC.1 shows the complete hospital survey questionnaire and possible responses. This survey was administered to patients via doctors and the results were translated from Bengali to English.

| Question                                                                 | Response Type          |
|-------------------------------------------------------------------------|------------------------|
| P1. What was the approximate time of the emergency?                     | Free text              |
| P2. When did you decide to leave for the hospital?                      | Free text              |
| P3. Which ward did you leave from?                                      | Ward Number            |
| P4. What time did you leave?                                            | Free text              |
| P5. What was your method of transportation?                             | A. Own car             |
|                                                                         | B. Rental car          |
|                                                                         | C. Rickshaw            |
|                                                                         | D. Ambulance           |
|                                                                         | E. CNG                 |
|                                                                         | F. Taxicab             |
|                                                                         | G. Other               |
| P6. What was the cost of transportation (in BDT)                        | A. <100                |
|                                                                         | B. 100-500             |
|                                                                         | C. 500-1000            |
|                                                                         | D. 1000+               |
| P7. Do you have a mobile phone?                                         | A. Yes                 |
|                                                                         | B. No                  |
| P8. Do you know how to contact an ambulance?                            | A. Yes                 |
|                                                                         | B. No                  |
| P9. Why did you not take an ambulance?                                  | Free text              |
| P10. Why did you come to this hospital?                                 | Free text              |
| H1. Name of hospital                                                    | Free text              |
| H2. What was the arrival time of the patient?                           | Free text              |
| H3. What is the general type of injury/complication?                    | Free text              |
| H4. What time did the patient first receive treatment?                  | Free text              |
EC.2. Demand data analysis

In this section, we conduct additional descriptive data analysis relevant to our project. Figure EC.1 shows a map of Dhaka’s 92 wards including those with a surveyed hospital and Table EC.2 provides summary statistics for key census characteristics. As noted above, ambulance usage was found to be quite low. Aside from cost, another possible explanation for low ambulance utilization is that ambulances were one of the slowest modes of transportation, while rickshaws, private cars, and other, which includes walking, were the fastest modes of transportation. Figure EC.2 displays the distribution of travel time from the trip origin to hospital for each mode of transportation. However, these results require careful interpretation. First, there is an extremely large variation in travel time for each mode, ranging from less than 5 minutes to over 8 hours. Second, many patients use rickshaws (and private cars) for shorter trips as shown in Table EC.3. Higher rickshaw usage, especially for short trips, is likely because rickshaws are readily available at all times and at nearly any location.

EC.3. Demand Estimation

In this section, we detail our methodology for estimating emergency transport demand. EC.3.1 focuses on population estimation, EC.3.2 focuses on estimating the expected number of ED visits per capita, and EC.3.3 presents our machine learning approach to predict the proportion of ED visits arriving via mode $M$. 

Figure EC.1  The map of Dhaka showing all 92 wards including wards without hospitals, wards with hospitals, and wards with surveyed hospitals.
Table EC.2  Demographic summary statistics across Dhaka’s 92 wards.

| Characteristic                        | Dhaka (all wards) | Individual ward | Minimum | Mean  | Maximum |
|---------------------------------------|-------------------|-----------------|---------|-------|---------|
| Population                            | 7,349,324         |                 | 18,170  | 79,884| 228,870 |
| Average household size                 | 4.3               |                 | 3.0     | 4.4   | 5.3     |
| Male-female population ratio           | 1.2               |                 | 0.9     | 1.3   | 2.5     |
| Population under 19 (%)               | 36.2              |                 | 22.9    | 35.5  | 43.5    |
| Population over 60 (%)                | 4.5               |                 | 2.4     | 4.5   | 7.6     |
| Married (%)                           | 59.0              |                 | 29.4    | 57.4  | 66.7    |
| Literacy (%)                          | 73.7              |                 | 52.5    | 74.9  | 90.4    |
| Pukka* house (%)                      | 58.4              |                 | 24.9    | 66.4  | 96.5    |
| Jupri* house (%)                      | 2.1               |                 | 0       | 1.8   | 11.2    |
| Sanitary toilet (%)                   | 58.0              |                 | 7.4     | 60.3  | 98.1    |
| Electricity (%)                       | 98.4              |                 | 92.5    | 98.8  | 99.9    |
| Rent-free home (%)                    | 3.4               |                 | 0.4     | 3.4   | 14.7    |
| Male-female employment ratio          | 2.0               |                 | 0.3     | 2.8   | 12.8    |

*Note that a Pukka house is a solid permanent dwelling usually made from brick or stone that is reflective of higher socioeconomic status, while a Jupri house is a temporary dwelling typically made from tin and other available supplies.

Figure EC.2  Distribution of travel time by mode of transportation.

EC.3.1. Estimating daytime ($n_{DT}^{w}$) and nighttime ($n_{NT}^{w}$) population

We obtained the daytime population in each ward from the Earthquake Vulnerability Assessment of Dhaka, which was conducted by the Government of Bangladesh with support from the United Nations Development Programme. In 2008, the total daytime population in Dhaka was estimated to be 6.63 million people. The nighttime population in each ward is obtained directly from the 2011 census and was estimated to be 7.35 million people. The population of Dhaka has been consistently growing at a rate of approximately 320,000 people per year [Streatfield and Karar 2008]. Under
the assumption that each ward is growing at a rate proportional to its population, we estimate the
total 2016 daytime and nighttime population in Dhaka to be 8.24 and 8.95 million, respectively.

**EC.3.2. Estimating the expected annual number of ED visits per capita ($\lambda_w$)**

In this section, we estimate the expected annual number of ED visits per capita in ward $w$ by
estimating the total number of annual ED visits in Dhaka. A recent study of a ED arrivals at a
“specialty corporate hospital” in Dhaka, found an average of 10,000 ED visits each year [Karim
et al. 2009]. This result requires careful interpretation because specialty hospitals can be very
expensive and serve only a limited population. To the best of our knowledge, there is no further
data on ED visits in Dhaka or other cities in Bangladesh.

To supplement this lack of data, we estimate the number of ED visits using data from other
similar South Asian cities. A study of three major government hospitals in Kirachi, Pakistan found
an average of 70,000 – 100,000 annual ED visits per hospital [Raftery 1996]. A similar study of
two major hospitals in New Dehli, India found that a private hospital with free emergency services
received 30,000 annual ED visits, while a government funded hospital with free services received
over 100,000 annual ED visits [PoSaw et al. 1998].

From these reports, we estimate the number of annual ED visits for a government funded hospi-
tal to be between 70,000 – 100,000 and we estimate the number of annual ED visits for a private
hospital to be between 10,000 – 30,000. Dhaka has 94 hospitals with EDs, of which 19 are govern-
ment funded. From this information, we estimate the number of annual ED visits to be between
2.08 – 4.15 million. Given that Dhaka has a population of 8.95 million, the visit rate is between
230 – 460 per 1000 persons. Therefore, the average number of annual ED visits per capita, $\lambda_w$, is
estimated to be between 0.23 – 0.46.

It is difficult to put these numbers into context because most LMIC countries do not collect
data on annual ED visits. However, data is available for 19 high income OECD countries and the

### Table EC.3

A breakdown of the inter-ward distance travelled by each mode of transportation.

| Mode       | Trips within ward (%) | Median travel distance for all trips (m) | Median travel distance for out of ward trips (m) |
|------------|-----------------------|-----------------------------------------|-----------------------------------------------|
| Rickshaw   | 30.2                  | 1358                                    | 1544                                          |
| CNG        | 6.8                   | 4749                                    | 5233                                          |
| Rental Car | 7.8                   | 6041                                    | 7018                                          |
| Private Car| 41.8                  | 1367                                    | 2462                                          |
| Ambulance  | 9.2                   | 3379                                    | 3670                                          |
| Taxi       | 5.0                   | 5737                                    | 6262                                          |
| Other      | 24.0                  | 2017                                    | 4041                                          |
average number of annual ED visits per capita across all countries is 0.31 with a range from 0.07 to 0.70 (Berchet 2015). These results require careful interpretation because they are from high income countries and they combine data from both rural and urban areas, which are known to have significant differences in ED visit rates. For example, in the US, urban areas have an annual rate of 0.32 ED visits per capita as compared to 0.45 ED visits per capita in rural areas. The only reliable data available for large urban areas is from New York City and Shanghai, which are similar to Dhaka in terms of population, but not in terms of culture or demographics. The annual per capita ED visit rates in New York City and Shanghai are 0.37 (Goins and Conroy 2015) and 0.33 (Zhang et al. 2014), respectively.

**EC.3.3. Estimating the proportion of ED visits arriving via mode \( M \) \( (\delta^M_w) \)**

In this section, we fit a regularized logistic regression model to predict the proportion of ED visits arriving via mode \( M \). In particular, we aim to estimate the proportion of ED visits arriving via ambulance \( (\delta^A_w) \) and the proportion of ED visits arriving via CNG \( (\delta^C_w) \), for each ward in Dhaka, respectively.

The nature of our data presents significant challenges for model training. Our survey data indicates that only 74 of Dhaka’s 92 wards include at least one surveyed patient and as a result, we cannot directly estimate \( \delta^M_w \) for each ward. Furthermore, only 32 of 92 wards include at least 20 observations. Given that the overall ambulance usage \( (\delta_A) \) is 7%, roughly 20 observations are required to ensure sufficient granularity in our estimations. One way to overcome the lack of data is to assume that \( \delta^M_w \) is uniform across all 92 wards and estimate a single proportion, \( \delta^M \), by grouping all wards together. Intuitively, assuming that \( \delta^M \) is uniform across all wards is analogous to using population as a proxy for the annual number of ambulance trips. We use this naive approach as a benchmark for our models. A second approach to overcome the lack of data is to group patients according to the ward in which the trip originated and calculate \( \delta^M_w \) for each ward with 20 or more patients (i.e., only 32 wards). We employ this approach and weight each \( \delta^M_w \) by the number of observations (i.e., patients) from ward \( w \). As a result, our final dataset includes 32 ward observations comprising 1,843 patients. For each ward, the features are a set of demographic variables selected from the census and the target is the proportion of ED trips via mode \( M \), i.e., \( \delta^M_w \), weighted according to the number of observations from ward \( w \).

Grouping patients by ward as opposed to a patient level approach that treats each patient as a unique observation is beneficial for our application for three key reasons: 1) we are interested in estimating \( \delta^M_w \) at the ward level, not the patient level, 2) we want our approach to be generalizable and this framework allows other regions in LMICs with only census data to apply our models, and 3) we require independent variables or features that are available for all 92 wards. The only features
available to us for all 92 wards are from the census and we link this data to the survey data (and mode choice) using the ward where the trip originated. In contrast, a patient level approach will cause all patients from the same ward to have identical independent variables, regardless of their mode choice.

The set of 27 demographic features we use in each model was selected from the census data, which contains 104 unique fields. To do this, we first remove all highly correlated ($R^2 > 0.85$) variables that appear to represent the same latent feature. In particular, we remove the minimum number of variables required to eliminate all pairwise correlations above 0.85. Next, we combined variables to create new features that have been previously shown to correlate with ambulance demand. For example, the original data contained male population, female population, and total population, which are all highly correlated. We kept total population and created a new variable using the ratio of male to female population. After this procedure, a final set of 27 demographic census features remained.

Our data is well suited for logistic regression because our observations can be viewed as independent Bernoulli trials and modelled using a binomial distribution. Given the large set of features and the likelihood of overfitting, we consider a logistic regression model with L1-regularization (LASSO), where $\gamma$ denotes the regularization parameter. We optimize over 1000 values of $\gamma$ between 0.0001 and 0.01. Recall, that the naive prediction approach mentioned above predicts a constant equal to the average $\delta^M$ across all wards. We also consider a weighed naive approach that predicts a constant equal to the weighted average $\delta^M$ across all wards.

We train our models using repeated 10-fold cross validation, which partitions the data into ten sets: eight sets of three wards and two sets of four wards. Each set is used exactly once as the testing set, while the remaining 9 are combined and used as the training set. We repeat this process 500 times to reduce the variance in our estimations of model accuracy. We measure prediction accuracy using root mean squared error

$$RMSE(y, \hat{y}) = \frac{1}{n}||y - \hat{y}||_2,$$

where $y \in \mathbb{R}^n$ are the true targets, $\hat{y} \in \mathbb{R}^n$ are the predicted targets, and $n$ is the number of samples. Once the value of $\gamma$ that minimizes RMSE is determined through repeated cross validation, we train a final model using this $\gamma$ and all available data to estimate $\delta^M_w$ for all 92 wards. All models were implemented using R version 3.3.3.

Figure [EC.3] displays a box plot of the RMSE distribution across the 500 repetitions for ambulance transport ($\delta^A_w$). The solid black line indicates the median and the box indicates the interquartile range. The whiskers extend to 1.5 times the interquartile range. The median RMSE was 0.04385
and 0.03935 for the naive and weighted naive, respectively. The logistic regression model performed the best with a median RMSE of 0.03886, corresponding to a 11.4% improvement over the naive approach and a 1.2% improvement over the weighted naive approach. Both improvements were found to be statistically significant at $\alpha = 0.05$ using the Wilcoxon signed-rank test. Although the logistic regression model only marginally improves upon the weighted naive approach, the model is able to provide insight into demographic features that contribute to ambulance usage as outlined below (also, see Table EC.4).

The logistic regression model with $\lambda = 0.00183$ was selected as the best method and trained using all available data. Table EC.4 provides information regarding the non-zero regression coefficients from the final model. The selected features include measures of population (avg. household size and ratio of males to females), a measure of social status (female marriage rate), and a measure of economic status (access to electricity). These results echo findings from previous research in developed urban centers in North America. In particular, we find that areas of higher social and economic status are more likely to use ambulances. For example, two of the wards with $\delta_{w}^{A} > 0.1$ include areas with a high density of foreigners and government officials and an area with major government offices, hospitals, and universities.

Figure EC.4 displays a box plot of the RMSE distribution across the 500 repetitions for CNG transport ($\delta_{w}^{C}$). The median RMSE was 0.180 and 0.181 for the naive and weighted naive, respectively. The logistic regression model preformed the best with a median RMSE of 0.176, corresponding to a 2.2% improvement over the naive approach and a 2.8% improvement over the weighted naive approach. Both improvements were found to be statistically significant at $\alpha = 0.05$ using the Wilcoxon signed-rank test.
Table EC.4  Non-zero regression coefficients as determined by LASSO for $\delta^A_w$.

| Feature                                      | Coefficient |
|----------------------------------------------|-------------|
| Intercept                                    | -4.065      |
| Average household size (number of persons)   | -0.085      |
| Ratio of male to female population           | 1.358       |
| Female marriage rate (%)                     | -0.014      |
| Access to electricity                        | 0.005       |

Figure EC.4  Comparison of RMSE between the logistic regression model and the naive approaches for $\delta^C_w$.

The logistic regression model with $\lambda = 0.0068$ was selected as the best method and trained using all available data. Table EC.5 provides information regarding the non-zero regression coefficients from the final model. We observe an increase in both the number and diversity of the regression coefficients for predicting $\delta^C_w$ as compared to $\delta^A_w$.

Although we focused on predicting the probability that a patient chooses an ambulance or CNG given that they require transportation to an ED (i.e., $\delta^M_w$, where $M = \{A, C\}$), our approach can be readily adapted to focus on other modes of transportation, such as rickshaws or motorcycles.

**EC.4. Travel time analysis**

**EC.4.1. Travel time data**

Figures EC5(a) and EC5(b) display the number of data samples per sampled edge and the number of edges per trip in the ambulance road network, respectively.
Table EC.5  Non-zero regression coefficients as determined by LASSO for $\delta^C_w$.

| Feature                                      | Coefficient |
|----------------------------------------------|-------------|
| Intercept                                    | -13.600     |
| Ratio of male to female population           | 0.062       |
| Male marriage rate (%)                       | -0.002      |
| Population between 0-19 (%)                  | 0.0077      |
| Population over 60 (%)                       | -0.149      |
| Disability rate (%)                          | -0.390      |
| Pukka house (%)                              | -0.028      |
| Access to a sanitary toilet with seal (%)    | 0.003       |
| Access to electricity (%)                    | 0.126       |
| Ratio of male to female employment           | 0.019       |

Figure EC.5  Histograms for collected GPS data. Only sampled edges are included in these figures.

EC.4.2. Travel time prediction

We compared the accuracy of four popular machine learning models: AdaBoost, Random Forest, linear regression with L1-regularization (LASSO), and K-nearest neighbors (KNN). For AdaBoost, we optimize the learning rate over $\{0.0001, 0.001, 0.01, 0.1, 1\}$ and number of weak learners over $\{100, 250, 500, 750, 1000\}$; for Random forest, we optimize the number of trees over $\{100, 250, 500, 750, 1000\}$; for LASSO, we optimize the regularization parameter over $\{0.0001, 0.001, 0.01, 0.1, 1\}$; for KNN we optimize the number of neighbors over $\{100, 250, 500, 750, 1000\}$. Each model contained an internal cross-validation procedure for hyper-parameter tuning. We train our models using repeated 10-fold cross validation and we repeat the cross validation process 100 times. We measure prediction accuracy using root mean squared error (RMSE). Once the final hyperparameters are determined, we train the model using all available data to obtain the final predictions of travel time, which are used in our optimization model. All
experiments were implemented using Python 3.5. Figures EC.6 and EC.7 display boxplots of the root mean squared error (RMSE) for each of the prediction models tested on edge and trip data, respectively.

To quantify the impact of geographical census features on prediction accuracy, we trained our sub-trip models with only distance and time features. Figure EC.8 displays a boxplot comparing the RMSE of our models trained with all features and with only distance and time features. The Lasso, AdaBoost, and RandomForest models improved when geographic features were included, and these improvements were found to be statistically significant using the Wilcoxon signed-rank test. The KNN model did not improve when geographical features were added because, unlike the other three models, KNN does not have an internal feature weighting process. In other words, the KNN model values all 73 features equally. By using only the distance and time features, we are implicitly selecting the most important features for the model. It may be possible that a small set of carefully selected geographical features added to the distance and time features will further improve KNN performance.
Figure EC.8 Mean-squared error results for sub-trip models with only distance and time features and with all features.

EC.5. Proof of Theorem 1.

Proof (by construction). Note that without vector notation, we can re-write NFF as:

\[
\min_{y, f} \sum_{(i,j) \in E} c_{ij} f_{ij}
\]

subject to

\[
\begin{align*}
\sum_{i \in N} y_i &= P, \\
\sum_{j \in O(i)} f_{ij} - \sum_{j \in I(i)} f_{ji} &\leq \alpha_i y_i - d_i, \forall i \in N, \\
y_i &\in \{0,1\}, \forall i \in N,
\end{align*}
\]

where \( I(i) = \{j \in N | (j,i) \in E\} \) and \( O(i) = \{j \in N | (i,j) \in E\} \). Recall the classic \( p \)-median formulation. The facility location variable is defined as \( x_{ii} = 1 \) if a facility is located at node \( i \in N \) and the assignment (routing) decision variables is denoted \( x_{ij} = 1 \) if demand node \( j \) has been assigned to facility node \( i \). Using this notation, the \( p \)-median problem (ReVelle and Sain 1970) can be formulated as:

\[
\min_{x} \sum_{i \in N} \sum_{j \in N} d_{ij} x_{ij}
\]

subject to

\[
\begin{align*}
\sum_{i \in N} x_{ii} &= P, \\
\sum_{i \in N} x_{ij} &= 1, \forall j \in N, \\
x_{ii} &\geq x_{ij}, \forall i, j, i \neq j, \in N, \\
x_{ij} &\geq 0, \forall i, j \in N, i \neq j, \\
x_{ii} &\in \{0,1\}, \forall i \in N,
\end{align*}
\]

where \( t_{ij} \) denotes the shortest travel time between nodes \( i, j \in N \) (\( i \) and \( j \) need not be adjacent) and \( t_{ii} = 0, \forall i \in N \). The optimal solution to (EC.2) is denoted by \( \hat{x} \).
We first show that the $p$-median is polynomially reducible to NFF. That is, we show that the optimal solution from the $p$-median can be transformed into the optimal solution of NFF in polynomial time and both solutions have the same optimal cost. First, set $\bar{y}_i = \bar{x}_{ii}, \forall i \in N$. By definition, the demand weighted shortest path length from $w \in N$ to $r \in N$ is given by $d_r t_{wr}$. To find the path from $w$ to $r$ (i.e., the sequence of nodes $i_w, ..., i_r$) along which flow must be directed, we solve:

$$\begin{align*}
\text{minimize} & \quad \sum_{(i,j) \in E} c_{ij} f_{ij}^r \\
\text{subject to} & \quad \sum_{i \in O(r)} f_{ri}^w - \sum_{i \in I(r)} f_{ir}^w = d_w, \\
& \quad \sum_{j \in O(i)} f_{ij}^w - \sum_{j \in I(i)} f_{ji}^w = 0, \forall i \in N \setminus \{r, w\}, \\
& \quad \sum_{j \in O(w)} f_{wj}^w - \sum_{i \in I(w)} f_{iw}^w = -d_w, \\
& \quad f_{ij} \geq 0, \forall (i, j) \in E.
\end{align*}$$

(EC.3)

We denote the optimal solution to (EC.3) as $\hat{\mathbf{y}}, \hat{\mathbf{f}}$. For the special case when $r = w$, Formulation (EC.3) is not well-defined and we assume that the shortest path has length zero (i.e., no flow is produced and $f_{ij}^w = f_{ij}^r = 0, \forall i, j \in N$). Set $\tilde{f}_{ij} = \sum_{r \in N} \sum_{w \in N} \tilde{f}_{ij}^r \tilde{x}_{rw}$ to obtain a solution $(\tilde{\mathbf{y}}, \tilde{\mathbf{f}})$ to Formulation (EC.1). We now show that the obtained solution $(\tilde{\mathbf{y}}, \tilde{\mathbf{f}})$ is feasible with respect to (EC.1).

For the first constraint, we have:

$$\sum_{i \in N} \tilde{y}_i = \sum_{i \in N} \tilde{x}_{ii} = P.$$

For the second constraint, define $J = \{j \in N \mid y_j = 0\}$ and $I = \{i \in N \mid y_i = 1\}$. Note that $I \cup J = N$. Consider some $k \in J$ (i.e., $y_k = 0$),

$$\sum_{j \in O(k)} \tilde{f}_{kj} - \sum_{j \in I(k)} \tilde{f}_{jk} = \sum_{j \in O(k)} \sum_{r \in N} \sum_{w \in N} \tilde{f}_{kj}^w \tilde{x}_{rw} - \sum_{j \in I(k)} \sum_{r \in N} \sum_{w \in N} \tilde{f}_{jk}^w \tilde{x}_{rw},$$

$$= \sum_{r \in N} \sum_{w \in N} \tilde{x}_{rw} \left( \sum_{j \in O(k)} \tilde{f}_{kj}^w - \sum_{j \in I(k)} \tilde{f}_{jk}^w \right),$$

$$= \sum_{r \in N \setminus \{k\}} \tilde{x}_{rk} \left( \sum_{j \in O(k)} \tilde{f}_{kj}^r - \sum_{j \in I(k)} \tilde{f}_{jk}^r \right) + \sum_{w \in N \setminus \{k\}} \tilde{x}_{kw} \left( \sum_{j \in O(k)} \tilde{f}_{kj}^w - \sum_{j \in I(k)} \tilde{f}_{jk}^w \right),$$

$$+ \sum_{r \in N \setminus \{k\}} \sum_{w \in N \setminus \{k\}} \tilde{x}_{rw} \left( \sum_{j \in O(k)} \tilde{f}_{kj}^w - \sum_{j \in I(k)} \tilde{f}_{jk}^w \right) + \tilde{x}_{kk} \left( \sum_{j \in O(k)} \tilde{f}_{kj}^k - \sum_{j \in I(k)} \tilde{f}_{jk}^k \right),$$

$$= \sum_{r \in N \setminus \{k\}} \tilde{x}_{rk} (-d_k) + \sum_{w \in N \setminus \{k\}} \tilde{x}_{kw} (d_w) + \sum_{r \in N \setminus \{k\}} \sum_{w \in N \setminus \{k\}} \tilde{x}_{rw}(0) + \tilde{x}_{kk}(0),$$

$$= -d_k.$$
Consider some \( k \in I \) (i.e., \( y_k = 1 \)),
\[
\sum_{j \in O(k)} \hat{f}_{kj} - \sum_{j \in I(k)} \tilde{f}_{jk} = \sum_{j \in O(k)} \sum_{r \in N} \sum_{w \in N} \hat{f}_{rw} \hat{x}_{rw} - \sum_{j \in I(k)} \sum_{r \in N} \sum_{w \in N} \tilde{f}_{rw} \tilde{x}_{rw},
\]
\[
= \sum_{r \in N} \sum_{w \in N} \hat{x}_{rw} \left( \sum_{j \in O(k)} \hat{f}_{kj} - \sum_{j \in I(k)} \tilde{f}_{jk} \right),
\]
\[
= \sum_{r \in N \setminus \{k\}} \hat{x}_{rk} \left( \sum_{j \in O(k)} \hat{f}_{kj} - \sum_{j \in I(k)} \tilde{f}_{jk} \right) + \sum_{w \in N \setminus \{k\}} \hat{x}_{kw} \left( \sum_{j \in O(k)} \hat{f}_{kj} - \sum_{j \in I(k)} \tilde{f}_{jk} \right)
+ \sum_{r \in N \setminus \{k\}} \sum_{w \in N \setminus \{k\}} \hat{x}_{rw} \left( \sum_{j \in O(k)} \hat{f}_{rw} - \sum_{j \in I(k)} \tilde{f}_{rw} \right) + \hat{x}_{kk} \left( \sum_{j \in O(k)} \hat{f}_{kj} - \sum_{j \in I(k)} \tilde{f}_{jk} \right)
\]
\[
= \sum_{w \in N \setminus \{k\}} \hat{x}_{rw} \left( \sum_{j \in O(k)} \hat{f}_{rw} - \sum_{j \in I(k)} \tilde{f}_{rw} \right) + \hat{x}_{kw} \left( \sum_{j \in O(k)} \hat{f}_{rw} - \sum_{j \in I(k)} \tilde{f}_{rw} \right)
\]
\[
\leq \sum_{w \in N \setminus \{k\}} d_w = \sum_{w \in N} d_w - d_k = \alpha - d_k.
\]

Lastly, we show that the objective function values of both solutions are equal,
\[
\sum_{(i, j) \in E} c_{ij} \hat{f}_{ij} = \sum_{(i, j) \in E} c_{ij} \sum_{r \in N} \sum_{w \in N} \hat{f}_{rw} \hat{x}_{rw},
\]
\[
= \sum_{r \in N} \sum_{w \in N} \hat{x}_{rw} \sum_{(i, j) \in E} c_{ij} \hat{f}_{rw},
\]
\[
= \sum_{r \in N} \sum_{w \in N} \hat{x}_{rw} d_w t_{rw}.
\]

We now prove the reverse direction. That is, a solution from NFF can be transformed into a solution for the \( p \)-median with the same optimal cost. We denote the optimal solution to NFF as \((\hat{y}, \hat{f})\).

First, we set \( \hat{x}_{kk} = \hat{y}_k, \forall k \in N \). Define \( J = \{ r \in N \mid \hat{y}_r = 0 \} \) and \( I = \{ w \in N \mid \hat{y}_w = 1 \} \). Note that \( I \cup J = N \). Compute \( t_{rw}, \forall r \in I \) and \( \forall w \in J \) (i.e., the shortest path between nodes \( r \) and \( w \)). This can be done by using Dijkstra’s algorithm or by extracting the path lengths directly from the given optimal solution to NFF. Both methods are polynomial time.

Now, consider some \( k \in J \), and solve \( \text{argmin}_{i \in I} t_{ik} \). Denote the optimal index as \( i^k \) and the optimal value as \( t_{ik} \). Set \( \hat{x}_{jk} = 1, \tilde{x}_{kj} = 0, \forall j \in N \), and \( \hat{x}_{ik} = 0, \forall i \in N \setminus \{i^k\} \). Consider some \( k \in I \), which implies that \( \hat{x}_{kk} = 1 \). Set \( \hat{x}_{ik} = 0, \forall i \in N \setminus \{k\} \) to obtain the solution, \( \hat{x} \). We now show that the obtained solution \( \hat{x} \) is feasible for the \( p \)-median.

For the first constraint, we have:
\[
\sum_{i \in N} \hat{x}_{ii} = \sum_{i \in N} \hat{y}_i = P.
\]
For the second constraint, consider \( r \in J \). By our construction, \( \tilde{x}_{ir} = 1 \) and \( \tilde{x}_{ir} = 0, \forall i \in N \setminus \{ i' \} \).

Therefore, \( \sum_{i \in N} \tilde{x}_{ir} = \tilde{x}_{ir} + \sum_{i \in N \setminus \{ i' \}} \tilde{x}_{ir} = 1 \). Consider, \( w \in I \). By our construction, \( \tilde{x}_{iw} = 1 \) and \( \tilde{x}_{iw} = 0, \forall i \in N \setminus \{ w \} \). Therefore, \( \sum_{i \in N} \tilde{x}_{iw} = \tilde{x}_{iw} + \sum_{i \in N \setminus \{ w \}} \tilde{x}_{iw} = 1 \). Combining these implies \( \sum_{i \in N} \tilde{x}_{ij} = 1, \forall j \in N \).

For the third constraint, consider \( r \in J \). By our construction \( \tilde{x}_{rr} = 0 \) and \( \tilde{x}_{rk} = 0, \forall k \in N \setminus \{ r \} \).

Therefore, \( \tilde{x}_{rr} \geq \tilde{x}_{rk}, \forall r, k \in N \setminus \{ r \} \). Consider \( w \in I \). By our construction, \( \tilde{x}_{ww} = 1 \) and \( \tilde{x} \in \{ 0,1 \} \) (i.e., \( \tilde{x} \leq 1 \)), therefore we have \( \tilde{x}_{ww} \geq \tilde{x}_{wk}, \forall w, k \in N \setminus \{ w \} \). Combining these implies \( \tilde{x}_{ii} \geq \tilde{x}_{ij}, \forall i, j \in N \).

Lastly, we show that the objective function values are equal. First we must derive some intermediate information. Consider the following optimization problem with \( \hat{y} \) fixed,

\[
\begin{align*}
\text{minimize} & \quad \sum_{(i,j) \in E} f_{ij} c_{ij} \\
\text{subject to} & \quad \sum_{j \in O(i)} f_{ij} - \sum_{j \in I(i)} f_{ji} \leq \alpha \hat{y}_i - d_i, \forall i \in N, \\
& \quad f_{ij} \geq 0, \forall (i,j) \in E.
\end{align*}
\]

Denote the optimal solution of \((\text{EC.4})\) by \( \hat{f} \). The dual of \((\text{EC.4})\) is given by,

\[
\begin{align*}
\text{maximize} & \quad \sum_{i \in N} p_i (\alpha \hat{y}_i - d_i) \\
\text{subject to} & \quad p_i - p_j \leq c_{ij}, \forall (i,j) \in E, \\
& \quad p_i \leq 0, \forall i \in N.
\end{align*}
\]

Denote the optimal solution to \((\text{EC.5})\) by \( \hat{p} \). The dual variable \( \hat{p}_k \) represents the change in optimal cost due to increasing \( d_k \) by one unit. If we increase \( d_k \) by one unit, the optimal solution will increase by the length of the shortest path from \( i \in I \) to \( k \). Therefore, at optimality, the value of \(-p_k\) (because \( p_k \) is negative in \((\text{EC.5})\)) is equal to the length of the shortest path from \( i \in I \) to \( k \in J \).

Mathematically, \(-p_k = t_{ik,k} \). Note that this implies that \( p_k = 0, \forall k \in I \) because the shortest path from a facility to itself, has length zero.
Now we show that optimal costs are equal:

\[
\sum_{r \in R} \sum_{w \in N} \tilde{x}_{rw} d_w t_{rw} = \sum_{r \in I} \sum_{w \in N} \tilde{x}_{rw} d_w t_{rw} + \sum_{r \in I} \sum_{w \in (r)} \tilde{x}_{rw} d_w t_{rw} + \sum_{w \in J} \tilde{x}_{rw} d_w t_{rw} + \tilde{x}_{ww} d_w t_{ww}
\]

\[
= \sum_{r \in I} \sum_{w \in N} \tilde{x}_{rw} d_w t_{rw} \quad (t_{ww} = 0, \tilde{x}_{rw} = 0, \forall r \in J, w \in N, \text{ and } x_{wr} = 0, \forall w \in I, r \in I \setminus \{w\})
\]

\[
= \sum_{w \in J} d_w \sum_{r \in I} \tilde{x}_{rw} t_{rw} \quad \text{(By our construction)}
\]

\[
= - \sum_{w \in J} d_w \hat{p}_w \quad \text{(From duality)}
\]

\[
= - \sum_{w \in N} d_w \hat{p}_w \quad (\hat{p}_w = 0, \forall w \in J)
\]

\[
= \alpha \sum_{w \in N} y_w \hat{p}_w - \sum_{w \in N} d_w \hat{p}_w \quad (\hat{p}_w = 0, \forall w \in I \text{ and } y_w = 0, \forall w \in J)
\]

\[
= \sum_{w \in N} \hat{p}_w (\alpha y_w - d_w)
\]

\[
= \sum_{(r,w) \in E} \hat{f}_{rw} c_{rw} \quad \text{(By strong duality)}
\]

Combining both directions, we have that

\[
\sum_{(i,j) \in E} \bar{f}_{ij} c_{ij} = \sum_{i \in I} \sum_{j \in N} \bar{x}_{ij} d_j t_{ij} \leq \sum_{i \in I} \sum_{j \in N} \tilde{x}_{ij} d_j t_{ij}, \forall \tilde{x},
\]

and that

\[
\sum_{i \in I} \sum_{j \in N} \bar{x}_{ij} d_j t_{ij} = \sum_{(i,j) \in E} \bar{f}_{ij} c_{ij} \leq \sum_{(i,j) \in E} \tilde{f}_{ij} c_{ij}, \forall \bar{f}.
\]

Therefore,

\[
\sum_{i \in I} \sum_{j \in N} \bar{x}_{ij} d_j t_{ij} = \sum_{(i,j) \in E} \tilde{f}_{ij} c_{ij}. \quad \square
\]

**EC.6. Proof of Theorem 2.**

*Proof.* Let \( Y = \{ y \mid e^t y = P, y \geq 0 \} \) and \( F(y, d) = \{ f \mid Af \leq \alpha I y - d, f \geq 0 \} \). Then, R-NFF can be written as

\[
\min_{y \in Y} \max_{c \in C, d \in D} \min_{f \in F(y, d)} c' f,
\]

or in epigraph form

\[
\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad t \geq \max_{c \in C, d \in D} \min_{f \in F(y, d)} c' f.
\end{align*}
\]
Enumerating the elements of $D$, the model becomes

$$\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad t \geq \max_{c \in C} \min_{f \in F(y,d^k)} c'f, \quad k = 1, \ldots, N.
\end{align*}$$

Since $C$ and $F(y,d^k)$ are disjoint, we can swap the min and max using the min-max theorem (Neumann 1928):

$$\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad t \geq \min_{f \in F(y,d^k)} \max_{c \in C} c'f, \quad k = 1, \ldots, N.
\end{align*}$$

We then replace $f$ by $f^k$ for each scenario $k$, which yields:

$$\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad t \geq \min_{f^k \in F(y,d^k)} \max_{c \in C} c'f^k, \quad k = 1, \ldots, N.
\end{align*}$$

We can now move $f^k$ to the outer minimization problem:

$$\begin{align*}
\text{minimize} & \quad t \\
\text{subject to} & \quad t \geq \maximize_{c \in C} c'f^k, \quad k = 1, \ldots, N, \\
& \quad f^k \in F(y,d^k), \quad k = 1, \ldots, N.
\end{align*}$$

Finally, for each $k$, we take the dual of the inner maximization problem to obtain the required result. □

EC.7. Comparison of solution approaches

Figure EC.9 displays the performance of HSGen as a function of the number of random starts and random interchanges for different numbers of scenarios and different numbers of outposts. Figure 9(a) shows that HSGen improves significantly from one to ten random starts, but does not appear to improve much beyond ten. By contrast, Figure 9(b) displays a small improvement from one to ten random starts and only marginal improvements thereafter. Thus, we use ten random starts to conduct our real experiments on the Dhaka road network. Figures 9(c) and 9(d) show that there does not appear to be a correlation between increasing the number of random interchanges and the overall solution quality. However, in all cases, there is a small improvement from one to 10 random interchanges. Thus, we use we use ten random interchanges to conduct our real experiments on the Dhaka road network.

EC.8. Dhaka Experiments

Figure EC.10 depicts the locations of all 67 current outposts.
EC.8.1. What is impact of the travel time budget?

In this section, we conduct a sensitivity analysis to determine the impact of the travel time budget. We use the HSGen algorithm with 10 random starts and 10 random interchanges to solve (6) with $P = 20$ and $B = \{0, 100, 1000, 2500, 5000, 7500, 10000\}$. We then apply the outpost locations resulting from a budget of 1000 seconds to all seven budget instances. We compare these results with the response time of outpost locations optimized and evaluated on the same budget. We conduct separate experiments for each of the three temporal combinations.

Figure EC.11 compares the response time performance between a fixed travel time budget and a problem-specific travel time budget for each of the three temporal combinations. For rush hour, the fixed budget outpost locations perform better when evaluated on networks with budgets of 2500 and 10000 with an average improvement of 1.45 minutes (4.5%). For all other instances, the
EC.8.1.1. Discussion and policy implications. Our results suggest that the outpost locations determined using a travel time budget of 1000 seconds are relatively insensitive to changes the travel time budget. This is an important result that implies that ambulance providers in Dhaka can use the optimal outpost locations from a budget of 1000 seconds without concern that these locations will perform significantly worse for other travel time budgets.

EC.8.2. What is the value of re-positioning current outpost locations?

Figure EC.11 displays the distribution of average response time for each number of re-positioned outposts during weekend and overnight. Similar to rush hour, re-locating a single outpost reduces response times by 1.5 minutes (15.0%) and 1.8 minutes (13.6%) for overnight and weekend, respec-
Re-positioning all 67 outposts provides overall response time improvements of 5.7 minutes (57%) and 7.6 minutes (58%) for overnight and weekend, respectively.

Figure EC.13(a) depicts the location of the old and new outpost for overnight and weekend. The road network travel time between the old and new locations for overnight and weekend is 80.8 and 80.2 minutes, respectively. Although the old outpost locations are different, the new outpost location is the same for both overnight and weekend. The new location chosen for both overnight and weekend is also very near to the new outpost location chosen for rush hour. These results further demonstrate that this area is undeserved for all temporal combinations.

Figures EC.13(b), EC.13(c), and EC.13(d) display the location of all current outposts, the old outpost locations, and the new outpost locations when re-positioning all 67 outposts for rush hour, overnight, and weekend. During rush hour, 16 outposts remain the same and the average travel time distance between an old and new outpost is 44.0 minutes. In contrast, only 7 and 10 outpost remain the same for overnight and weekend, respectively. The average distance between an old and new outpost for overnight and weekend is 10.1 and 12.4 minutes, respectively. Across all three instances, a significant number of new outposts are located in the bottom left, bottom center, and bottom right wards suggesting that these areas are under-served. In particular, the bottom center area with a cluster of road network nodes includes Old Dhaka - a densely populated area with many narrow streets.

EC.8.3. What is the value of adding additional outpost locations to the current network?

Figure EC.14 displays the response time distribution for each number of additional outposts. Adding a single new outpost for both overnight and weekend provides the largest value with reductions of 1.5 minutes (15.0%) and 1.5 minutes (11.5%), respectively. Figure EC.15 displays the locations of 20 new outposts for overnight and weekend. The addition of 20 new outposts provides a reduction in response time of 4.8 minutes (47.8%) and 5.9 minutes (44.6%) for overnight and
(a) A single repositioned outpost for overnight
(b) All 67 outposts repositioned during rush hour and weekend

(c) All 67 outposts repositioned during overnight
(d) All 67 outposts repositioned during weekend

Figure EC.13 The location of new and old repositioned outposts.

Figure EC.14 Response time performance of adding additional outpost locations.

weekend, respectively. These results suggest that adding 20 new outposts provides similar value to re-positioning all 67 outposts.
EC.8.4. What is the value of designing a completely new emergency response network?

Figure EC.16 displays the distribution of average response time for overnight and weekend. The response time performance of 20 new outpost locations is 0.5 minutes (5.0%) and 0.5 minutes (3.6%) worse than the current 67 outposts for overnight and weekend, respectively. Figure EC.17 depicts the location of 20 new outposts. The are five current outposts that are the same as the new locations for both overnight and weekend, respectively. In general, the overnight and weekend locations are quite similar to each other with 9 of the same locations. On the other hand, the rush hour locations are typically quite different from the overnight and weekend locations with only 5 and 2 outposts in common, respectively.
Figure EC.17  The locations of 20 new outposts.