Select, Route and Schedule: Optimizing Community Paramedicine Service Delivery with Mandatory Visits and Patient Prioritization

Shima Azizi[1], Özge Aygül[2], Brenton Faber[3], Sharon Johnson[4], Renata Konrad[4], Andrew C. Trapp[2,4]
[1]Business Analytics and Information Systems, Peter J. Tobin College of Business, St. John's University, Queens, NY, USA
[2]Data Science Program, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA USA
[3]Biomedical Engineering, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA USA
[4]The Business School, Worcester Polytechnic Institute, 100 Institute Road, Worcester, MA USA

{azizis@stjohns.edu, oaygul@wpi.edu, bdfaber@wpi.edu, sharon@wpi.edu, rkonrad@wpi.edu, atrapp@wpi.edu}

Abstract: Healthcare delivery in the United States has been characterized as overly reactive and dependent on emergency department care for safety net coverage, with opportunity for improvement around discharge planning and high readmissions and emergency department bounce-back rates. Community paramedicine is a recent healthcare innovation that enables proactive visitation of patients at home, often shortly after emergency department and hospital discharge. We establish the first optimization-based framework to study efficiencies in the management and operation of a community paramedicine program. The collective innovations of our modeling include i) a novel hierarchical objective function with the goals of fairly increasing patient welfare, lowering hospital costs, and reducing readmissions and emergency department visits, ii) a new constraint set that ensures priority same-day visits for emergent patients, and iii) a further extension of our model to determine the minimum supplemental resources necessary to ensure feasibility in a single optimization formulation. Our medical-need based objective function prioritizes patients based on their clinical features and seeks to select and schedule patient visits and route healthcare providers to maximize overall patient welfare while favoring shorter tours. We use our methods to develop managerial insights via computational experiments on a variety of test instances based on real data from a hospital system in Upstate New York. We are able to identify optimal and nearly optimal tours that efficiently select, route, and schedule patients in reasonable timeframes. Our results lead to insights that can support managerial decisions about establishing (and improving existing) community paramedicine programs.

Keywords: Community Paramedicine, Integer Optimization, Team Orienteering, Mandatory Visits, Time Windows.

1. Introduction

Healthcare in the United States has been criticized for being fragmented, poorly-integrated, and reactive. Emergency departments (ED) stand-in for primary care especially for patients who rely on safety net coverage or require after-hours care. Discharge planning after inpatient care remains tenuous as evidenced by high readmissions rates across the country. It has been estimated that more than 18 billion USD could be saved annually if patients whose medical problems are considered non-urgent were to take advantage of primary or preventive health care rather than relying upon the ED for their medical needs [1]. These lower-acuity patients are at risk of becoming increasingly ill and eventually requiring acute ED care and potentially a costly hospital admission. Further, the cost of unplanned readmissions post-discharge has been estimated to be 15 to 20 billion USD annually [2]. Hospitals in the United States face penalties for 30-day readmissions and avoidable ambulatory care sensitive conditions [3] and actively seek ways to prevent both inpatient readmissions and ED bounce-backs.

Community paramedicine (CP) is a recent healthcare delivery innovation that, in contrast to a reactive approach, enables *proactive* visitation of patients in the comfort of their own home, often shortly after an ED or hospital discharge. CP has been shown to be effective for chronic disease and ED high-utilizer management and, in some cases, for acute conditions that could be treated at home rather than at an ED. Patients recently discharged from the hospital often require follow-up care such as medication administration and adherence, wound management, or routine check-ups. Other complaints such as dehydration, nausea, flu, constipation, or anxiety do not necessarily require ED level resources. Without appropriate resources, these recently-discharged and ill patients are at greater risk for unplanned readmission or ambulatory sensitive ED visit. Moreover, while ED resources should be used for true emergencies, many patients use ED services for non-urgent purposes. CP aims to increase access to primary and preventive care and decrease unnecessary ED visitation and unplanned hospital readmission, ultimately seeking to increase patient welfare and decrease health care costs [4, 5].

CP, alternatively called *mobile integrated care* uses traditional front-line clinicians to help bridge gaps between primary care and community health needs. Prehospital care in the United States typically involves three levels of certification, Emergency Medical Technician (EMT, Basic Life Support), Paramedic (Advanced Life Support), and Intermediate Emergency Medical Technician (jurisdiction-dependent hybrid between EMT and Paramedic). While CP programs utilize all three levels of specialty, going forward we will use the term "community paramedic" for efficiency while acknowledging that some services may be best provided by other levels of care [6].

Unlike ambulance and transport-based paramedicine, community paramedics work where patients live (homes, shelters, assisted living centers) to prevent their health conditions from deteriorating to the point where emergency and inpatient care are required. Whereas emergency paramedics are an in-the-field extension of an emergency medicine (EM) physician, the community paramedic can be an extension of either EM or a primary care (PC) physician.

In the United States, a community paramedic's training, scope-of-practice, and credentialing is determined by each state. While some states and more rural and under-resourced jurisdictions have found benefit in expanding the paramedic's scope of practice to include a broader range of medications and diagnostic practices (e.g., phlebotomy, urinalysis reagent tests, wound care), other jurisdictions maintain the paramedic's usual scope but expand where and in what ways the paramedic can practice and how agencies can be compensated for these services [6]. The International Board of Speciality Certification offers Community Paramedic certification alongside other specialty certifications for paramedics (Flight, Critical Care, Tactical, Tactical Responder, Wilderness) [7]. In most jurisdictions, certification remains optional as CP is still in development, is rapidly evolving, and has yet to be widely adopted [8].

CP shares some similarities to home health care (HHC) services. While both programs provide health-focused services performed at client homes, most home health care services are typically assessment-driven and can operate on a relatively fixed and routine schedule. CP programs, in contrast, are designed to bring certain emergency department-level services to the home by expanding the role of advanced paramedics and emergency medical technicians (EMTs) who are routed through the community. In CP, community paramedics fill gaps in community care and complement existing services provided by HHC, visiting nurses, and social workers. In this way, the community paramedic works best as part of a multidisciplinary patient care team that can include case managers, visiting nurses, behavioral health specialists, social workers, and physicians [9].

Important clinical differences exist between CP and HHC and these differences have implications for scheduling and operations. Whereas nurses and other HHC providers have expertise in longer term patient care, as an extension of an emergency medicine or primary care physician, the community paramedic may be able to provide in-home interventions not available to usual HHC providers. In addition to patient assessments, CP personnel perform interventions in the home including medication administration, IV therapy, EKGs, phlebotomy, urine tests, disease screens (such as rapid COVID-19), and, in some jurisdictions, vaccinations. Community paramedics typically also use a designated vehicle (fly car[1] or personal car) whereas cars, bikes, public transport and walking are used in HHC. Since HHC typically focuses on assessing patients and ensuring that specific clinical goals are being met (medication adherence, weight loss, medication reconciliation post-discharge), two critical differences exist between delivering CP and HHC services.

First is the requirement of who must be visited in a given planning period and second is the amount of time required for each visit. First, in the CP delivery model, higher-severity patients must be visited as these patients require clinical interventions that are necessary to prevent an inpatient or emergency-department encounter. Visits to lower-severity patients could be moved to another day if postponing that visit would not result in a more expensive clinical encounter. Thus, the priority for CP is to visit as many emergent patients as possible within a given time period (such as a day), while lower-severity patients, if not selected to be visited, may be postponed and placed on less-frequent schedules. Second, CP visits are more dynamic than HHC as CP interventions can take as little as 15 minutes or as long as two hours. In the HHC model, if a patient requires advanced care (IV fluids for example), emergency services such as 911 are dispatched and the patient is sent to the emergency department. In the CP model, if a patient requires advanced care, the visiting clinician provides that care after consultation with a physician. In this way, scheduling CP services requires a more dynamic approach that aligns patient severity and the requirements for in-home treatment.

To the best of our knowledge, this study establishes the first optimization formulation to solve the challenges of CP selection, routing, and scheduling, valid over any planning period. Routing and scheduling challenges for mobile and HHC services have seen exposure in recent literature, as further detailed in Section 2; however, the authors are only aware of manual approaches to tackle CP delivery challenges. We use optimization-based approaches to transform perhaps the most challenging part of the CP puzzle: replacing the complex, manual decision-making process of selecting patients – routing healthcare providers to visit patients and scheduling CP visits while respecting patient time windows.

Given existing resources and a set of patients to be visited, our optimization-based framework answers the following questions concerning the efficient management of a CP program: *1) Who are the priority-weighted patients that must be visited, by which vehicle, and in what order? 2)*

[1]A fly car is a minimally equipped emergency medical service vehicle that is not designed to transport patients.

How should CP resources be most efficiently allocated and routed? 3) Are there some patients who must be seen? 4) If there are insufficient resources available to visit all patients who must be seen, what level of supplemental resources are necessary to do so? We anticipate that answering these questions will reveal critical insights into efficiently managing a CP program. In addition to revealing the most efficient ways to select patients and schedule and route medical workers to maximize patient and hospital benefits, our framework informs the number of patients that can be treated in a program, the type of patients that should be visited, and the number of vehicles and personnel required to initiate a new program, or equivalently, are needed to sustain the program through acquisition or outsourcing.

Our optimization formulation makes several novel contributions to the HHC operations research literature. These include the introduction of a new objective function to maximize overall patient welfare, while secondarily minimizing the total distance traveled by the healthcare providers. Among ways to do so, we incorporate means to eliminate unnecessary distance traveled through a subsidiary objective function component. Moreover, we construct a unique way to consider fairness in selecting patients via a prioritization scheme according to associated clinical features. This novelty ensures that visiting emergent patients are prioritized.

Our model also accommodates the mandatory nature of visiting emergent patients, in that we require certain patient categories (patients with mandatory visits) to be visited using the available resources. Should existing resources prove insufficient, our model also prescribes the remaining supplemental resources that are necessary to accommodate the needs of these emergent patients. Moreover, we carefully structure our novel optimization formulation to ensure supplemental resources are used only when the existing resources are insufficient to meet the needs of the emergent patients, while ensuring the supplemental resources travel no more than necessary. Lastly, we present a thorough computational analysis using real data from our problem context, discuss results, and provide meaningful managerial insights.

The remainder of this study is organized as follows. In Section 2, we survey the related literature, focusing on existing CP pilot programs in the United States, as well as routing and scheduling of HHC services. In Section 3, we establish a new mixed integer linear formulation for optimizing CP service delivery and extend our baseline model to address further challenges that may occur in a CP program. In Section 4, we introduce our case study and its characteristics, conduct computational experiments on de-identified data from our partner organization to assess the validity of our mathematical formulations, and discuss the results of our computational experiments and insights. We draw conclusions and discuss future research directions in Section 5.

2. Literature Review

Within the operations research literature, the authors are unaware of any studies to model the selection, routing, and scheduling of visits for a CP program. We first survey the existing CP pilot programs in the United States, followed by a discussion of the analytical approaches used for related problems in HHC routing and scheduling literature.

2.1 Community Paramedicine Pilot Programs in the United States

In many countries including the United States, ED resources are limited by space, equipment, personnel, and budget [10]. At the same time, demands on ED services are frequent, and include interactions that can be categorized as *true emergencies* such as trauma, heart attacks, strokes, major organ dysfunction, or systemic infections (sepsis); *lower acuity conditions* such as disease complications (COPD, diabetes), wound care and minor trauma, ambulatory-sensitive conditions;

bouncebacks such as ED high utilizers; and *readmissions* such as recently discharged patients with complications that risk an unplanned readmission. It is important to reserve limited and expensive ED resources to quickly respond to true emergencies, as service delays for such patients may be life threatening [11]. Community paramedicine holds promise as a cost-effective healthcare delivery innovation, enabling healthcare systems to make better use of their limited resources by visiting potential bounceback, readmission, and lower-acuity patients proactively outside of the ED and hospital, while conserving limited ED resources for true emergencies.

In the United States, CP programs have been explored only recently, piloting in at least 20 states as of 2016 [12][2]. Characteristics and outcomes of several recent CP implementations in the United States are summarized in Table 1. These pilot programs differ in targeted patients and treatment offered at the patient home. Targeted patients are commonly frequent users of the ED, chronically ill patients, elderly populations, hospice patients, and patients with non-acute conditions, immunization needs, or patients at high-risk of unplanned hospital readmissions. CP treatments commonly offered include wound care and chronic disease management, primary care, follow-up care after hospital and ED discharge, and educating patients on the difference between urgent care and primary care [12].

Several pilot programs attempt to evaluate CP impacts in terms of reducing healthcare provider costs and improving societal health. Between 2010 and 2015, MedStar Mobile Healthcare in Dallas and Fort Worth, Texas, claims to have prevented a total of 1.893 ED transports for 146 patients, saving Medicare more than 800 million USD [12]. An initial evaluation of a small pilot program in Abbeville County, South Carolina, found measurable improvements in patient health. In particular, a key finding was that many patients previously in need of consistent services only required occasional check-ups [4]. Another program in Eagle County, Colorado, reported saving 124.071 USD in healthcare costs over a two-year implementation [4]. While such examples illustrate the impact of CP, the emphasis of this study is on determining how best to route and schedule patients after target populations and services are defined, rather than evaluating the impact of CP programs.

The CP concept originated in rural settings as a strategy to increase access to basic healthcare needs. More recently, it has found increased use in more urban areas like Albuquerque, New Mexico [13]. Community paramedicine is a part of the Emergency Medical Service Agenda 2050 outlining the future version of emergency medical service in the United States [14]. At the same time, the 2050 Agenda underscores the importance of using analytical approaches to design and implement CP programs. In the next section, we investigate the HHC problem which, from an operations research viewpoint, is the most related context to CP.

[2] Community paramedicine pilots may be more widespread since 2016.

Table 1: Examples of Community Paramedicine Pilot Program in the United States.

Location	Proactive Treatment	Targeted Patients	Reported Benefits
Livingston County, NY [15]	Trained EMTs treat geriatrics at home	Geriatric	None reported
Madison,WI [12]	Paramedics provide wound care & chronic disease management	Chronically ill patients	None reported
Clayton County, GA [12]	Paramedics treat patients at home	ED high utilizers: $\geq$ 17 visits/year	None reported
Dallas and Fort Worth, TX [12]	Educate & monitor patients	Chronically ill patients	ED visits: 1893 (↓) Medicare saving: > $800 million
Albuquerque, NM [13]	Visit & educate patients	ED superusers & patients at high risk for readmission	Frequent users' utilization of ED: 70% (↓)
Rugby, N.D. [16]	Paramedics & EMTs bring medical care at home	Chronically ill & hospice patients	ED & hospital admissions: (↓) Patient satisfaction: (↑)
Abbeville County, SC [4]	Visit patient at home	Frequent ED users: $\geq$ 1 chronic disease	ED visits: 58.7% (↓) Inpatient stays: 68.8% (↓) 30-day readmission rates: 41.2% (↓)
Eagle County, CO [4]	Paramedics treat patients at home	Patients with non-acute condition or immunization need	Healthcare cost saving: $124,071 Average saving per visit: $1,969
Central Jackson County, MO [17]	Paramedics followed up with patients for a year	Congestive heart failure patients	Readmission rate: 15% (↓)

2.2 Scheduling and Routing of Home Health Care Services

CP is a unique operational model that shares some similarities with HHC. We now survey related works in the HHC domain from an operations research perspective and highlight the features that distinguish CP from other health delivery models in the literature. The use of operations research approaches to address and mitigate HHC challenges is a relatively new and challenging area of study, and considers issues such as caregiver-to-patient assignment, scheduling of patient requests, and caregiver routing [18, 19]. Such factors create a complex decision-making environment which is difficult to address in a rigorous and formal way, and results in the individual planning aspects generally being considered separately in the literature.

While it is challenging to mathematically model and subsequently optimize all aspects of the HHC problem, a number of studies consider separate aspects in a mathematical model [18–26]. Eveborn et al. [23] formulate the combined scheduling and routing problem as a set partitioning problem. Cappanera and Scutellà [18] propose an integrated approach to jointly address assignment, scheduling, and routing decisions in the home care problem through a novel concept of patterns, which specifies a possible schedule for skilled visits. They use patterns as a key strategy to reduce the complexity. Clapper et al. [27] address the design of caregiver shift patterns, considering the timing and type of caregiver to be scheduled.

Nikzad et al. [28] propose a two-stage stochastic mixed integer model with uncertainty in travel and service times. They partition the set of patients and caregivers into districts in the first stage while considering operational costs, handling routing and assignment decisions in the second stage. Uncertainty in service and travel times has more recently been considered in the HHC domain. Di Mascolo et al. [29] provide a literature survey on routing and scheduling problems in HHC, highlighting prominent techniques such as stochastic, dynamic, and robust optimization. In stochastic modeling approaches, uncertain variables may follow known distributions such as normal [30–32] and uniform [33]. Relatedly, Hoot et al. [34] develop a discrete event simulation of an emergency department (ED) and emphasize that ED procedures differ in their distributions. They use the lognormal distribution for evaluation and treatment of patients in the ED.

Fikar and Hirsch [35] provide a review of the HHC scheduling and routing studies from an operations research perspective, dividing HHC scheduling and routing studies into single period and multi-period timeframes. Many HHC studies [20, 21, 23–25, 36] model the problem in a single-period, assuming a single working day as the planning horizon. In multi-period HHC studies [18, 37–41], service requests and availability of nurses may change over the planning horizon, with the planning period being defined as different days of a week or month.

Fikar and Hirsch [35] also categorize HHC scheduling and routing studies in terms of their objective functions. Common objective functions include minimizing travel time [21, 24, 36], travel cost [20, 23, 25], wait time [21], overtime [21, 36], as well as maximizing nurse preferences [20, 21, 24, 25, 36] and balancing workload among nurses [18, 19]. Alkaabneh and Diabat [42] propose two different algorithms that aim to minimize service and routing costs while maximizing compatibility between caregivers and patients by putting weights on the objective function. Recently, Cinar et al. [39] appear to be the first to consider maximizing the total priority of the visited patients primarily and minimizing the total traveling time secondarily in a single vehicle HHC problem. To the best of our knowledge, we are the first to do so in the context of multiple nonhomogeneous vehicles and differing patient priority scores: that is, consider maximizing overall patient welfare, while favoring shorter tours.

The vast majority of HHC studies focus on visiting all patients [18, 19, 43]. However, in CP settings, healthcare systems must prioritize their resources and often have insufficient capacity to visit all patients. Given insufficient resources to visit all patients, it becomes necessary to prioritize visiting patients with the greatest need, which we refer to as *emergent* patients. Typical HHC models of routing service providers, based on classical problems like the traveling salesman problem and variations of the vehicle routing problem, are not appropriate for CP, as they assume *all* patients must be visited, when in reality available resources require patient selection. We instead employ a team orienteering approach with time windows to model CP delivery challenges, which selects patients for visits, routes the healthcare providers to visit the selected patients, and in so doing prescribes schedules for the selected visits, while respecting patient time windows.

The classical vehicle routing problem aims to minimize either the number of vehicles serving all vertices or the total travel distance for a fixed number of vehicles, making it challenging to compare with the team orienteering problem due to distinct objectives. However, in the selective vehicle routing problem, not all vertices can be visited due to constraints such as limitations on vehicle tour length, and vehicle capacity for serving vertices with specific demands. The selective vehicle routing problem seeks to maximize the possible reward without violating any of its constraints. The selective vehicle routing problem and the team orienteering problem primarily differ in the type of constraints. The team orienteering problem has additional constraints based on attributes unique to each vertex (such as service time; entrance fee), while the distance constraint in the selective vehicle routing problem is based on distances between vertices and is therefore dependent on the visit sequence [44–47].

The orienteering problem is a combination of node selection and determining the shortest Hamiltonian path between the selected nodes. In the orienteering problem, each node has a certain score and the objective is to maximize the total collected score by visiting as many (priority-weighted) nodes as possible, whereas the traveling salesman problem attempts to minimize the cost (as measured by time, distance, or economic costs) to visit all nodes. Team orienteering is an extension of the orienteering problem that considers a specific number of agents conducting tours, with each respecting individual time limits [45]. Over the last decade, a number of challenging practical applications have been modeled using the team orienteering framework [45]. However, Cinar et al. [39] is the only study that uses the *orienteering* viewpoint in the HHC literature, and we are the first to model CP delivery challenges via the *team orienteering* viewpoint. Furthermore, only a few exact approaches [18–20, 39] have been proposed that deal simultaneously with the various planning aspects in HHC. Cinar et al. [39] exactly solve instances of moderate size, but to generate solutions for larger problems they resort to an Adaptive Large Neighborhood Search algorithm and a matheuristic.

In real life situations such as CP, the limited availability of resources makes visiting all sites potentially impractical. In such settings, a number of important questions arise that can be addressed through proper mathematical modeling and analysis: *1) Who are the priority-weighted patients that should be visited, by which vehicle, and in what order? 2) How should community paramedicine resources be most efficiently allocated and routed? 3) Are there some patients that must be seen? 4) If there are insufficient resources available to visit all patients who must be seen, what level of supplemental resources are necessary to do so?* Our optimization-based framework can answer all of these questions within a single optimization formulation, offering key insights and analytical support for medical decision makers.

While studies exist in the literature that address some aspects of our approach, none are able to simultaneously consider all of the elements we do. We consider selecting patients in addition to routing and scheduling of healthcare providers, as well as fairness in the selection decisions through the construction of a functional representation of the benefit of visiting each patient based on the associated health features. Our model is the first optimization formulation in the HHC literature to ensure treatment is provided for patients with severe health conditions by including an obligation to visit certain categories of patients during the planning period (such as those that are emergent). Our model also incorporates the ability to find routes that are not longer than necessary. Finally, given the limited resources available to visit patients coupled with the necessity that some patients must be visited, it may be necessary for healthcare systems to obtain additional resources (either through outsourcing, or increasing their own internal resources) to ensure emergent patients are visited – our model considers this aspect for the first time and determines the required amount of supplemental resources.

3. Methodological Developments

In this section, we address the core challenges of the CP service delivery system through mathematical modeling. We begin with a formal problem description, provide a baseline formulation for the problem, and extend the baseline model to solve realistic challenges that may occur in a CP program.

3.1 Mixed-Integer Optimization Model for Community Paramedicine

A formal problem description of CP begins with requiring vehicles to start and end their tour at pre-specified depots. We attribute to each patient a single value called the *priority score*, indicating the urgency of a CP visit for such patients based on their clinical assessment. While it is desirable to visit all patients, it is assumed that there are insufficient resources to do so. The goal is to visit as many (priority-weighted) patients as possible while respecting travel times, time windows, and time budgets. We assume there exists a set of patients who should be scheduled over a given planning period.

Sets. Let $\mathcal{P}$ represent the set of patients and their associated locations, indexed by i. The set of clinical features affecting a patient priority score is denoted by $\mathcal{F}$. The set of vehicles is denoted as $\mathcal{V}$, indexed by v. The set of all nodes, including the origin depot, patient locations and the destination depot, is denoted by $\mathcal{N}$, indexed by i. Where notationally convenient we assume an ordered set of nodes, patients and vehicles. We denote the origin and destination depots either as nodes corresponding to the indices $i = 0$ and $i = |N|$, or as nodes O and D, and use one or the other based on notational convenience.

Parameters. Denote the priority score for each patient $i \in \mathcal{P}$ as p_i, where the function $p : \mathcal{P} \times \mathcal{F} \mapsto \mathbb{R}_+^{|\mathcal{P}|}$ defines these priority scores, and where $p_i > p_j$ implies that patient i has a greater priority than patient j. The visit duration for node $i \in \mathcal{N}$ is d_i, the travel time from node $i \in \mathcal{N}$ to node $j \in \mathcal{N}$ is t_{ij} $(i \neq j)$, and the given time budget for vehicle $v \in \mathcal{V}$ is T_v. Let M be a large enough constant to upper bound the start time of servicing any node. Denote the lower and upper bounds on the start of service for node $i \in \mathcal{P}$ as l_i and u_i, respectively.

Variables. Let binary variable y_i^v equal 1 if patient $i \in \mathcal{P}$ is visited by vehicle $v \in \mathcal{V}$, and 0 otherwise. Let binary variable x_{ij}^v $(i \neq j)$ equal 1 if a visit to node $i \in \mathcal{N}$ is followed by a visit to node $j \in \mathcal{N}$ by vehicle $v \in \mathcal{V}$, and 0 otherwise. Denote the start of service for node $i \in \mathcal{N}$ by vehicle $v \in \mathcal{V}$ as nonnegative continuous variable s_i^v.

With the above notation, we formulate the following mixed-integer optimization model that maximizes the total priority score from visiting patients:

$$\text{maximize} \quad \sum_{v=1}^{|\mathcal{V}|} \sum_{i=1}^{|\mathcal{P}|} p_i y_i^v \tag{1a}$$

$$\text{subject to:} \quad \sum_{v=1}^{|\mathcal{V}|} \sum_{j=1}^{|\mathcal{N}|-1} x_{0,j}^v = \sum_{v=1}^{|\mathcal{V}|} \sum_{i=1}^{|\mathcal{N}|-1} x_{i,|\mathcal{N}|}^v, \tag{1b}$$

$$\sum_{v=1}^{|\mathcal{V}|} \sum_{j=1}^{|\mathcal{N}|-1} x_{0,j}^v \leq |\mathcal{V}|, \tag{1c}$$

$$\sum_{\substack{i=0,\\ i\neq k}}^{|\mathcal{N}|-1} x_{ik}^v = \sum_{\substack{j=1,\\ j\neq k}}^{|\mathcal{N}|} x_{kj}^v = y_k^v \quad \forall\, k \in \mathcal{P}, \forall\, v \in \mathcal{V}, \tag{1d}$$

$$\sum_{v=1}^{|\mathcal{V}|} y_i^v \leq 1 \quad \forall\, i \in \mathcal{P}, \tag{1e}$$

$$\sum_{i=0}^{|\mathcal{N}|-1} \sum_{\substack{j=1,\\ j\neq i}}^{|\mathcal{N}|} t_{ij} x_{ij}^v + \sum_{i=1}^{|\mathcal{P}|} d_i y_i^v \leq T_v \quad \forall\, v \in \mathcal{V}, \tag{1f}$$

$$s_i^v + t_{ij} + d_i - s_j^v \leq M(1 - x_{ij}^v) \quad \forall\, i, j \in \mathcal{N}(i \neq j), \forall\, v \in \mathcal{V}, \tag{1g}$$

$$l_i y_i^v \leq s_i^v \leq u_i y_i^v \quad \forall\, i \in \mathcal{P}, \forall\, v \in \mathcal{V}, \tag{1h}$$

$$s_i^v \geq 0 \quad \forall\, i \in \mathcal{P}, \forall\, v \in \mathcal{V}, \tag{1i}$$

$$x_{ij}^v \in \{0, 1\} \quad \forall\, i, j \in \mathcal{N}(i \neq j), \forall\, v \in \mathcal{V}, \tag{1j}$$

$$y_i^v \in \{0, 1\} \quad \forall\, i \in \mathcal{P}, \forall\, v \in \mathcal{V}. \tag{1k}$$

Constraint set (1b) ensures that all tours start and end at origin and destination depots, and constraint (1c) ensures that the number of tours does not exceed the maximum number of vehicles. Constraint set (1d) guarantees the connectivity of the path for each vehicle. Constraint set (1e) ensures that each patient is visited at most once, while constraint set (1f) ensures all travel and servicing adheres to time budget limitations. Constraint set (1g) determines the timeline of the path for each vehicle, effectively serving to eliminate subtours. Constraint set (1h) ensures that the start time of patient visits occurs within time windows. Variable domains appear in (1i)– (1k).

3.2 Distance Prioritization

We can prioritize shorter tours with the following objective function, which maximizes the total weighted priority score of visited patients while favoring shorter tours:

$$\text{maximize} \quad \sum_{v=1}^{|\mathcal{V}|} \sum_{i=1}^{|\mathcal{P}|} p_i y_i^v - \alpha \left(\sum_{v=1}^{|\mathcal{V}|} \sum_{i=0}^{|\mathcal{N}|-1} \sum_{\substack{j=1, \\ j \neq i}}^{|\mathcal{N}|} t_{ij} x_{ij}^v + \sum_{v=1}^{|\mathcal{V}|} \sum_{i=1}^{|\mathcal{P}|} d_i y_i^v \right) \tag{2a}$$

where $\alpha > 0$ is a small enough constant to ensure that the contribution of the second objective function component is always strictly less than the smallest benefit that could accrue from the first objective function component.

3.3 Ensuring Treatment of Emergent Patients

In reality, there may exist some patients that must be visited during any given planning period. Assume such a set of patients $\mathcal{H} \subseteq \mathcal{P}$ exists, indexed by i. These patients must be visited during the present planning period because of the severity of their health conditions. Constraint set (3b) enforces this condition. This extension of the model replaces constraint set (1e) with constraint set (3a) and constraint set (3b).

$$\sum_{v=1}^{|\mathcal{V}|} y_i^v \leq 1 \quad \forall\, i \in \mathcal{P} \setminus \mathcal{H}, \tag{3a}$$

$$\sum_{v=1}^{|\mathcal{V}|} y_i^v = 1 \quad \forall\, i \in \mathcal{H}. \tag{3b}$$

If the new set $\mathcal{H}$ is empty, constraint set (3b) is vacuous, and constraint set (3a) simply reduces to constraint set (1e). When $\mathcal{H}$ is nonempty, additional required resources may need to be determined for fulfilling the needs of these emergent patients, which we next investigate in Section 3.4.

3.4 Determining Supplemental Resources

Even in the best circumstances, a situation may occur in which there are insufficient resources to serve all emergent patients requiring a visit in a defined time period. For such circumstances, we propose to add a supplemental set of resources in the model to determine the additional level of vehicles and time necessary to ensure all emergent patients are visited within the budgeted time. This feature of our framework can inform decision-makers to better understand the tradeoffs around available and needed resources, for example the acquisition of additional resources or outsourcing. Denote the set of existing vehicles as $\mathcal{V}_1$, indexed by v; and denote the set of supplemental vehicles as $\mathcal{V}_2$, indexed by v. Denote the set of all vehicles including available and supplemental vehicles as $\mathcal{V}^+ = \mathcal{V}_1 \cup \mathcal{V}_2$, indexed by v. If emergent patients are not visited in the planning period via available vehicles $v \in \mathcal{V}_1$ due to limited time resources, they will be visited by supplemental vehicles $v \in \mathcal{V}_2$.

The benefit of adding supplemental vehicles $v \in \mathcal{V}_2$ transforms the problem from one of being infeasible, to both ensuring that a feasible solution exists, as well as simultaneously capturing the minimum amount of supplemental resources needed to accommodate all emergent patients.

The objective function is structured so as to prioritize the use of available resources first, and supplemental resources only when necessary to ensure feasibility. To ensure that the use of supplemental resources only occurs when the available vehicles are exhausted, supplemental vehicles $v \in \mathcal{V}_2$ are not considered in the first term of the objective function. With the new notation, we present a final mixed-integer optimization model that maximizes the total priority score of visited patients via existing vehicles $v \in \mathcal{V}_1$, while favoring shorter tours for both available and supplemental vehicles. A small penalty $\beta > 0$ is also added to the model to favor shorter tours for the supplemental vehicles $v \in \mathcal{V}_2$. Our final model is presented as follows.

$$\begin{aligned}
\text{maximize} \quad & \sum_{v=1}^{|\mathcal{V}_1|}\sum_{i=1}^{|\mathcal{P}|} p_i y_i^v - \alpha\left(\sum_{v=1}^{|\mathcal{V}_1|}\sum_{i=0}^{|\mathcal{N}|-1}\sum_{\substack{j=1,\\ j\neq i}}^{|\mathcal{N}|} t_{ij}x_{ij}^v + \sum_{v=1}^{|\mathcal{V}_1|}\sum_{i=1}^{|\mathcal{P}|} d_i y_i^v\right) \\
& - \beta\left(\sum_{v=1}^{|\mathcal{V}_2|}\sum_{i\in\mathcal{H}\cup\{O\}}\sum_{\substack{j\in\mathcal{H}\cup\{D\},\\ j\neq i}} t_{ij}x_{ij}^v + \sum_{v=1}^{|\mathcal{V}_2|}\sum_{i\in\mathcal{H}} d_i y_i^v\right) && \text{(4a)}\\
\text{subject to:} \quad & \sum_{v=1}^{|\mathcal{V}^+|}\sum_{j=1}^{|\mathcal{N}|-1} x_{0,j}^v = \sum_{v=1}^{|\mathcal{V}^+|}\sum_{i=1}^{|\mathcal{N}|-1} x_{i,|\mathcal{N}|}^v, && \text{(4b)}\\
& \sum_{v=1}^{|\mathcal{V}^+|}\sum_{j=1}^{|\mathcal{N}|-1} x_{0,j}^v \leq |\mathcal{V}^+|, && \text{(4c)}\\
& \sum_{\substack{i=0,\\ i\neq k}}^{|\mathcal{N}|-1} x_{ik}^v = \sum_{\substack{j=1,\\ j\neq k}}^{|\mathcal{N}|} x_{kj}^v = y_k^v \quad \forall\, k\in\mathcal{P}, \forall\, v\in\mathcal{V}^+, && \text{(4d)}\\
& \sum_{v=1}^{|\mathcal{V}_1|} y_i^v \leq 1 \quad \forall\, i\in\mathcal{P}\setminus\mathcal{H}, && \text{(4e)}\\
& \sum_{v=1}^{|\mathcal{V}^+|} y_i^v = 1 \quad \forall\, i\in\mathcal{H}, && \text{(4f)}\\
& \sum_{i=0}^{|\mathcal{N}|-1}\sum_{\substack{j=1,\\ j\neq i}}^{|\mathcal{N}|} t_{ij}x_{ij}^v + \sum_{i=1}^{|\mathcal{P}|} d_i y_i^v \leq T_v \quad \forall\, v\in\mathcal{V}^+, && \text{(4g)}\\
& s_i^v + t_{ij} + d_i - s_j^v \leq M(1-x_{ij}^v) \quad \forall\, i,j\in\mathcal{N}(i\neq j), \forall\, v\in\mathcal{V}^+, && \text{(4h)}\\
& l_i y_i^v \leq s_i^v \leq u_i y_i^v \quad \forall\, i\in\mathcal{P}, \forall\, v\in\mathcal{V}^+, && \text{(4i)}\\
& s_i^v \geq 0 \quad \forall\, i\in\mathcal{P}, \forall\, v\in\mathcal{V}^+, && \text{(4j)}\\
& x_{ij}^v \in \{0,1\} \quad \forall\, i,j\in\mathcal{N}(i\neq j), \forall\, v\in\mathcal{V}^+, && \text{(4k)}\\
& y_i^v \in \{0,1\} \quad \forall\, i\in\mathcal{P}, \forall\, v\in\mathcal{V}^+. && \text{(4l)}
\end{aligned}$$

In optimization formulation (4a)–(4l), the parameters α, β, and M need to be appropriately set so as to suitably affect the performance of the solver and model outcomes. An appropriate value for each of these parameters is introduced in the next section.

3.5 Model Enhancements

Optimization formulations (1a)–(1k) and (4a)–(4l) contain big-M terms that are used to determine timeline of each path. We derive a large enough value for M to upper bound the start of service for each node, supporting off-the-shelf solvers in coping with these formulations. We also calculate appropriate values for α and β used to prevent available and supplemental resources from traveling more than necessary. Finally, we propose a novel functional representation for priority scores, considering the notion of fairness to prioritize patients with more urgent needs over patients with less urgent needs.

Deriving A Sufficient Big-M Value for Timeline. A sufficient value for the big-M constant for each vehicle $v \in \mathcal{V}$ in constraints (4h) is defined as

$$M_v = T_v + \gamma, \quad \forall\, v \in \mathcal{V} \tag{5}$$

where γ is a small enough positive constant. A straightforward argument can be made that the start of service for any node $i \in \mathcal{N}$ cannot be larger than time budget of vehicle $v \in \mathcal{V}$ that may visit node i.

Deriving Sufficient α and β Values. The objective function (4a) represents a lexicographic objective function, with alpha and beta parameters set to hierarchically prioritize the three objective components. The primary objective aims to maximize the total weighted priority score of visited patients, while the secondary objective is concerned with finding shorter tours for both existing and supplemental vehicles. The secondary objective, when appropriately weighted as in Section 4, can be considered as a tie-breaker that selects among multiple optima of the primary objective. In this way, the goal of maximizing the total weighted priority score of visited patients is given preference over minimizing the total length of vehicle tours. Therefore, the primary objective does not necessarily select the highest priority patients. By doing so, more emphasis is placed on potential savings from the intervention, which involves reducing the total number of emergency department visits. In some cases, this may mean that certain patients, based on geography or population density, may be disadvantaged. However, this approach enables resources to be more effectively allocated, ultimately resulting in improved outcomes for the population overall. A sufficient value for α is

$$\alpha = \frac{\min\limits_{i \in \mathcal{P}} p_i - \epsilon}{\sum_{v=1}^{|\mathcal{V}_1|} T_v} \tag{6}$$

where ϵ is a small, positive constant.

The value of α should fulfill $\alpha \leq \frac{\min_{i \in \mathcal{P}} p_i}{\sum_{v=1}^{|\mathcal{V}_1|} \sum_{i=0}^{|\mathcal{N}|-1} \sum_{j=1}^{|\mathcal{N}|} t_{ij} x_{ij}^v + \sum_{v=1}^{|\mathcal{V}_1|} \sum_{i=1}^{|\mathcal{P}|} d_i y_i^v}$, and $\sum_{v=1}^{|\mathcal{V}_1|} T_v$ is a valid upper bound on the denominator of the right-hand side of this inequality. Subtracting ϵ from the numerator ensures that in the worst case, the first objective function component that maximizes collective priority scores remains the emphasis. The value of β is set to $\frac{1}{\sum_{v \in \mathcal{V}_2} T_v}$ to penalize longer tours for supplemental vehicles, consistent with the structure of the primary purpose of the objective function to maximize the total weighted priority score of visited patients by available vehicles.

Priority Score Calculation. There is a strict preference in the patient selection process for the class of patients that we want to prioritize. It is unfair to let patients who are emergent wait a long time or even wait until the next planning period simply because visiting patients with less urgent needs is less expensive in terms of travel distance. We construct a function to accommodate this notion of fairness and to prioritize patients with more urgent needs over patients with less urgent needs. To facilitate the ensuing discussion, we assume that there is a modest number of levels of patient urgency over which there is a weak ordering[3], so that some patients may have identical urgency.

[3] If there are more than a modest number of levels representing urgency, or even continuous values, this can still be accommodated by stratification into a smaller number of levels.

The importance of visiting patient i is controlled via priority score p_i in objective function (4a). To prioritize visits so that patients with more urgent needs are visited first, it is necessary to assign a large enough score to these patients so that visiting multiple patients with less urgent needs would not be more favorable in the objective function than visiting a single emergent patient. In equation (7), we derive an appropriate expression for p_i that ensures more critical patients have a higher visit priority in the model.

For each patient $i \in \mathcal{P}$, let $\mathcal{L}_i$ be the set of all patients $\ell \in \mathcal{P}$ with less urgent needs than patient i. Define p_i as

$$p_i = \sum_{\ell \in \mathcal{L}_i} p_\ell + \epsilon \tag{7}$$

where ϵ is a small, positive constant.

In the worst case scenario, there are enough resources to visit all patients with less urgent needs than patient i within a given time frame, and yet patient i is not visited due to limited available resources. The total benefit obtained from visiting patient $\ell \in \mathcal{L}_i$ is $\sum_{\ell \in \mathcal{L}_i} p_\ell$. Therefore, visiting patient i will be preferred only if p_i is greater than $\sum_{\ell \in \mathcal{L}_i} p_\ell$. Hence, $\sum_{\ell \in \mathcal{L}_i} p_\ell + \epsilon$ suffices. Therefore, calculating p_i for each patient $i \in \mathcal{P}$ according to equation (7), ensures that patients with more severe health conditions have higher visit priority.

In practice, a health system can establish the initial priority scores based on a combination of factors, not solely on the physician-indicated acuity level. These factors include: 1) *ED acuity*, the acuity level assigned to a patient upon arrival at the ED, which provides an initial indication of the patient's medical urgency; 2) *Assigned acuity on discharge*, which accounts for the acuity level assigned to a patient at the time of discharge, and may differ from the initial ED acuity, offering a more accurate representation of the patient condition after evaluation and treatment; 3) *Chief complaint*, the primary concern or symptom reported by the patient during triage, which plays a critical role in differentiating the needs of low acuity and emergent patients and identifying those who may most benefit from the CP visits; 4) *Net gain of home treatment*, representing the potential benefits of providing home treatment such as preventing ED or inpatient events, and assists in identifying patients who may experience significantly improved health outcomes and optimized resource utilization through home treatment. By considering these factors, the initial priority scores provide a comprehensive assessment of each patient's needs and the potential benefits of the CP visit.

4. Computational Experiments

In this section we conduct computational experiments on test instances inspired by real hospital system data, to study the performance of optimization formulation (4a)–(4l). We begin by introducing our case study and addressing characteristics of our test instances. We then present the results of our computational experiments and discuss managerial insights related to our findings. While we consider loose patient time windows in our experiments that are in accordance with the expressed preference of our case study organization, we also investigate the performance of our model with tighter patient time windows. We further report our findings with uncertain visit duration and explore the use of supplemental vehicles.

All experiments are run on parallel computing resources with 1 node and 8 cores. We use Gurobi Optimizer version 8.0.4 [48] with Python 3.6.5 interface for solving optimization formulation (4). We use default Gurobi parameter settings for all of our experiments, and set a time limit of 10 hours (wall-clock) for each instance. All pairwise distances between nodes in set $\mathcal{N}$ are calculated via the HERE Map API [49].

4.1 Community Paramedicine in Canton-Potsdam Hospital: A Case Study

Potsdam is a rural town in St. Lawrence County, New York, with a population of 14,901 according to the latest available 2020 census. Canton-Potsdam Hospital is the sole hospital serving the town and surrounding community within a 50 mile radius, encompassing approximately 25,000 people. We use data from Canton-Potsdam Hospital in our computational experiments to help assess the feasibility of implementing a rural CP program for their patient population. A challenge to managing and implementing CP for this hospital is that the large and sparse population must be served via the limited resources of the hospital or a local rescue squad. It is therefore necessary to find the most efficient way to select the patients who should be visited and treat them with the existing resources. We next show that our model can provide answers for Canton-Potsdam Hospital concerning the management, design and implementation of an efficient delivery system for CP.

We generate various test instances inspired by real data from Canton-Potsdam Hospital, and measure the performance of our optimization model by varying key parameters affecting the problem size. We create test instances by varying the number of patients in the panel, number of vehicles, and time budget for each vehicle, as detailed in Table 2. We assume an equal time budget of T_v for all available vehicles. Appropriate values for α, β, and M are calculated according to values introduced in Section 3.5.

Table 2: Characteristics of Instances Used in Computational Experiments.

Number of Patients	Number of Available Vehicles	Time Budget for Each Available Vehicle (Hours)
10	1	{2, 3, 4, 5, 6}
	2	{2, 3, 4}
20	1	{2, 3, 4, 5, 6, 7, 8}
	2	{2, 3, 4, 5, 6}
30	1	{4, 5, 6, 7, 8, 9, 10}
	2	{4, 5, 6, 7}
{40, 50}	{1, 2}	{4, 5, 6, 7, 8, 9, 10}
{60, 70, 80, 90, 100}	{1, 2, 3}	{6, 7, 8, 9, 10, 11, 12}

Our case study data is sorted by patient discharge times from the hospital. For a particular scenario with $|\mathcal{P}|$ patients, we select the first consecutive unique $|\mathcal{P}|$ patients starting from a particular date in January. We repeat this sampling procedure for a random start date in June and October. This sampling method generates three instances that we used for each scenario with $|\mathcal{P}|$ patients. It is worth noting that the patients in each instance may be from different days, as the main purpose of our experiments is testing the model for a wide range of patients in the panel.

Our model uses data that includes patient address, acuity level, estimated visit duration, and patient time windows. Patient acuity level, assigned by a physician, is available for each patient indicating a patient's clinical assessment; this is indicated by an integer number between one and five (from least to most severe). We consider acuity level as a feature affecting the priority score. Based on clinical expertise about patient needs and CP services, we assume specified visit durations for various acuity levels[4], as detailed in Table 3. In our case study, we assume that patients with acuity level one and two do not require CP visits. Thus, the panel only includes patients with acuity level three, four, and five. We further explore uncertain visit duration following truncated lognormal distribution, which will be discussed in the next section.

Table 3: Visit Duration for Patients with Varying Acuity Levels.

Acuity Level	Visit Duration (Minutes)
3	15
4	20
5	30

Table 3 indicates that there are three groups of patients who constitute the panel. Patients with acuity level of five have the highest urgency for visitation, while patients with an acuity level of three have the lowest urgency. We assume a CP program would require patients with the highest acuity level to be visited in the present planning period. Thus, the set of emergent patients[5] $\mathcal{H}$ consists of patients with acuity level five, and we designate vehicles $v \in \mathcal{V}_2$ that can only be used for visiting this group of patients. If patients with acuity level five are not visited via available vehicles $v \in \mathcal{V}_1$ due to limited time availability, they must be visited with vehicles $v \in \mathcal{V}_2$. This capability allows our model to flex so as to determine the supplemental resources necessary to visit these acuity level five patients, thereby preventing the model from becoming infeasible due to constraint set (3b).

While each patient has their individual acuity level, the determination of objective function coefficients is dictated by equation (7), in particular because we aim to prioritize the most emergent patients over those that are less emergent in optimization formulation (4). Given a unique panel of patients, each instance has its own objective coefficients (priority scores) based on the panel properties. For calculating priority scores, we assume that ϵ is equal to 1 in equation (7).

Section 4.2 reveals key managerial insights we derive from our computational experiments. Canton-Potsdam Hospital staff generally requires patients to be available for the entire duration of the day. We thus set the patient time windows to be at their widest so that the lower bound on starting service for any patient is at the beginning of the day, while the upper bound is at the end of the day. Even so, to understand the effect of stricter time windows on the computational performance of optimization formulation (4), we conduct a thorough study on various time window scenarios. We also explore the significance of uncertainty in visit duration by conducting experiments with stochastic visit duration.

4.2 Managerial Insights from Computational Experiments

The model is designed to provide insights in answering critical questions for managers that oversee CP programs. First, *given a set of priority-weighted patients needing services, as well as a set of vehicles with predetermined characteristics, who are the patients that should be visited, by which vehicle, and in what order, while ensuring that CP resources are most efficiently allocated and routed?* Another important question to managers, as expressed by our partnering hospital, is

[4]We evaluate the effect of considering longer visit durations on the model performance and results in Appendix C.

[5]The set of emergent patients $\mathcal{H}$ was introduced in Section 3.3.

whether it is possible to visit all emergent patients with available resources. Finally, if our model determines that the available resources are insufficient to visit all emergent patients, it also is able to answer two key, related questions: *i) What percentage of emergent patients are not visited by available resources? ii) What level of additional resource is needed to visit these patients?* In the short term, this information might be used for an existing program to explore outsourcing options. Over the longer term, such data might be used to provide justification for investment in additional vehicles. In planning a new program, running the model multiple times to reflect different scenarios can provide insight regarding necessary resources.

We address these questions by studying the computational results for a variety of scenarios created by varying numbers of patients, numbers of vehicles, and time budgets per vehicle, using the instances outlined in Table 2. We evaluate computational times and optimality gaps to demonstrate the feasibility of our model in multiple dynamic settings beyond the standard case study setting, in particular scenarios with longer visit durations, as well as scenarios with patient-specified time windows.

Results Assuming A Standard Work Day. We discuss computational results under a standard eight hour work day assumption. Table 4 presents results from running optimization formulation (4) for instances with patient discharge times in January, while the time budget for each available vehicle T_v is eight hours. Note the column titled *"Number of Vehicles"* represents all vehicles including available vehicle(s) and one supplemental vehicle. We assume that the set $\mathcal{V}_2$ of supplemental vehicles includes only one vehicle with a large enough time budget to visit all patients with the highest acuity level should these patients not be visited via available vehicles $v \in \mathcal{V}_1$ due to their limited time availability. We can therefore find supplemental time for visiting emergent patients who are not visited by available vehicles[6]. In Appendix B, we include experimental results for 20, 40, 60, 80, and 100 patients, varying the number of vehicles, and even time budget values per vehicle, omitting the detailed results with 10, 30, 50, 70, and 90 patients and odd time budget values, as these results are similar to the reported results.

In Table 4, over all instances that are not solved to global optimality within the 10 hour time limit, the optimality gap is no more than 5.2%, on average 1.46%. We also highlight the time to find the best incumbent solution. In instances with limited resources and a considerable number of emergent patients with acuity level of five–such as the instances with 100 patients and two vehicles–available and supplemental vehicles are used to visit only emergent patients.

These findings reveal the following insight: our model can identify the required amount of available resources (how many vehicles, how much time budget) needed to visit k% of the patients, for $k = 100$, or any $k \in [0, 100]$ if insufficient resources are present for $k = 100$.

[6] In Appendix A, we present Algorithm A1 to determine the minimum number of supplemental vehicles needed to visit emergent patients.

Table 4: Selected Results of Solving Formulation (4) for Varying Number of Patients and Vehicles, While Time Budget for Each Available Vehicle is Eight Hours; a "–" indicates reaching time limit prior to solving to optimality.

# Patients	# Vehicles	# Patients			Supplemental	Seen Patients via Vehicle(s)		Runtime	Time to Find Best	Optimality
		Acuity 5	Acuity 4	Acuity 3	Time (Hours)	Available	Supplemental	(Seconds)	Incumbent (Seconds)	Gap
10	2	5	1	4	0	10 (100.0%)	0 (0.0%)	2	1	0%
20	2	5	7	8	0	13 (65.0%)	0 (0.0%)	3	2	0%
30	2	7	11	12	0	13 (43.3%)	0 (0.0%)	2	2	0%
40	2	11	16	13	2.4	14 (35.0%)	2 (5.0%)	5	5	0%
	3				0	26 (65.0%)	0 (0.0%)	5,730	1,827	0.01%
50	2	14	21	15	3.16	12 (24.0%)	3 (6.0%)	10	9	0%
	3				0	28 (56.0%)	0 (0.0%)	–	577	1.28%
60	2	18	22	20	5.7	12 (20.0%)	6 (10.0%)	21	18	0%
	3				0	26 (43.3%)	0 (0.0%)	–	34,913	1.49%
	4				0	40 (66.7%)	0 (0.0%)	–	26,821	3.09%
70	2	21	27	22	7.46	12 (17.1%)	9 (12.9%)	118	84	0%
	3				3.09	26 (37.1%)	2 (2.9%)	–	21,544	2.45%
	4				0	40 (57.1%)	0 (0.0%)	–	6,860	4.07%
80	2	22	30	28	9.06	14 (17.5%)	11 (13.8%)	1,380	1,261	0%
	3				3.67	27 (33.8%)	3 (3.8%)	–	29,157	2.56%
	4				0	42 (52.5%)	0 (0.0%)	–	22,340	3.92%
90	2	27	32	31	12.52	13 (14.4%)	15 (16.7%)	3,472	1,411	0%
	3				7.0	27 (30.0%)	7 (7.8%)	–	26,379	1.99%
	4				0	37 (41.1%)	0 (0.0%)	–	9,014	5.20%
100	2	32	36	32	15.15	13 (13.0%)	19 (19.0%)	9,965	9,538	0.01%
	3				8.46	26 (26.0%)	9 (9.0%)	–	31,087	2.36%
	4				3.63	40 (40.0%)	3 (3.0%)	–	30,020	3.61%

Which Patients Should be Visited, by Which Vehicle, and in What Order? To illustrate that the optimization approach addresses the first managerial question regarding which patients are visited and the routes and resources used, we present sample results considering different time budgets. In Figure 1 we plot the optimal routes for an instance with 10 patients, a single available vehicle, and one supplemental vehicle, while varying vehicle time budgets over six values. The location of patients with acuity level of three, four, and five are depicted via green, blue, and red empty circles, respectively. A black diamond indicates the location of the Canton-Potsdam Hospital depot. A yellow tour highlights the optimal route for the available vehicle, and a pink tour shows the optimal route for the supplemental vehicle.

In Figure 1a, the available vehicle time budget is set to just two hours. The available vehicle visits one patient with acuity level of five and a patient with acuity level of four. In this case, a supplementary vehicle is required to visit all patients with acuity level of five. In Figure 1b, we increase the available vehicle time budget to three hours. As a result, two patients with an acuity level of five and one patient with an acuity level of four are visited via the available vehicle, and the hospital needs less supplemental resources to visit all emergent patients.

Comparing the subplots in Figure 1 from top left to bottom right reveals less need for the supplemental vehicle as the time budget for the available vehicle increases – the optimal constructed tours become more similar and more reasonable in shape. In observing the optimal routes in Figure 1, it becomes clear that patients with higher acuity level are more frequently visited, while the patients who are at lower risk are visited only if sufficient resources exist. The last panel, Figure 1f, demonstrates that if there is a seven hour time budget, nearly all patients are visited by the available resource.

These findings reveal the following insight: a tradeoff exists between the level of allowable resources and the performance of the optimization model in terms of visiting emergent, and other less severe patients. When resource levels are relatively low, only emergent patients are prioritized (as specified by the assumptions of the CP problem), whereas relatively more resources allow for additional emergent and even less severe patients to be visited. The existence of a model to inform decision making concerning the number of available vehicles and corresponding time budgets is vital to understanding these tradeoffs and their implications on improving the overall welfare of the served community.

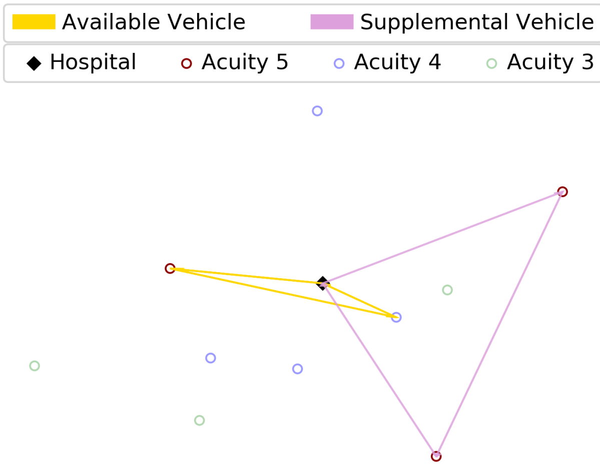
Who Are the Unseen Patients, and How Can Their Needs Be Accommodated? We consider the effect of varying time budgets on supplemental resource usage to answer the following key questions to managers: i) Is it possible to visit *all* emergent patients with available resources? ii) What percentage of emergent patients are *not* visited by available resources? iii) What level of additional resource (measured in time) is needed to visit these patients?

For instances with 50 patients, Figure 2 uses output from the optimal solutions to plot the average percentage of seen patients and the supplemental time necessary for visiting unseen emergent patients across increasing values of T_v, the time budget for each available vehicle $v \in \mathcal{V}_1$. For $|\mathcal{P}| = 50$, $\mathcal{V}^+ \in \{2, 3\}$, and $T_v \in \{4, 5, ..., 10\}$ for available vehicle $v \in \mathcal{V}_1$, we have three different instances of optimization formulation (4). Figure 2 depicts the average results over the three instances with $|\mathcal{P}| = 50$ patients, while varying the number of vehicles and the time budget. In Figure 2a, only one available vehicle exists for visiting patients, while Figure 2b demonstrates the results for two available vehicles. In both figures, the time budget in hours appears along the x-axis, against the left y-axis of average percent of seen patients, and the right y-axis of average supplemental necessary time, in hours, to visit unseen emergent patients.

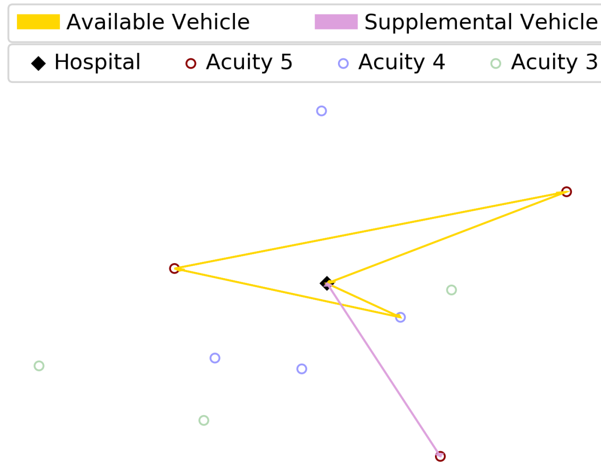
In Figure 2, each bar represents the average percentage of seen patients via available (green) and supplemental (blue) vehicles. For the runs with lower available vehicle time budgets, the average percentage of seen patients via the supplemental vehicle, and consequently the average supplemental

Figure 1: Comparing optimal routes for an instance with 10 patients, as the available vehicle time budget increases hour by hour.

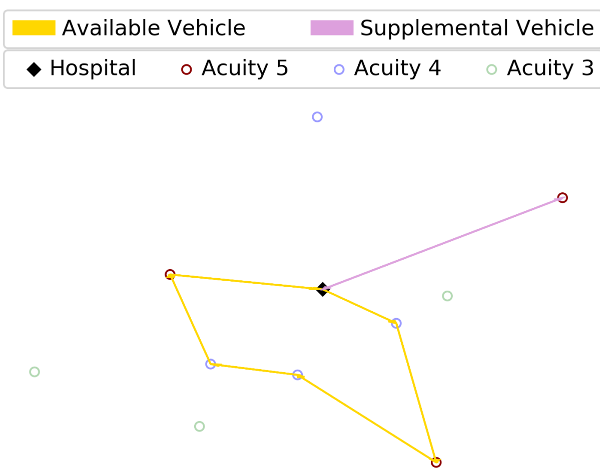
(a) Two hour time budget.



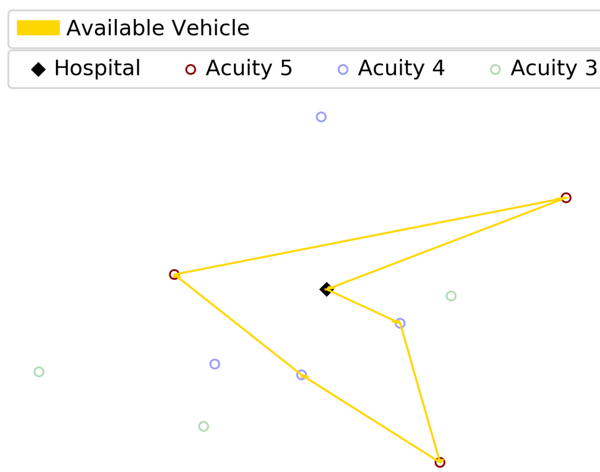
(b) Three hour time budget.



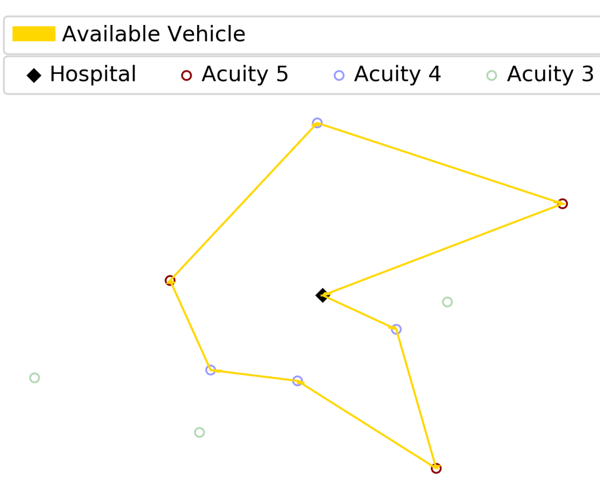
(c) Four hour time budget.



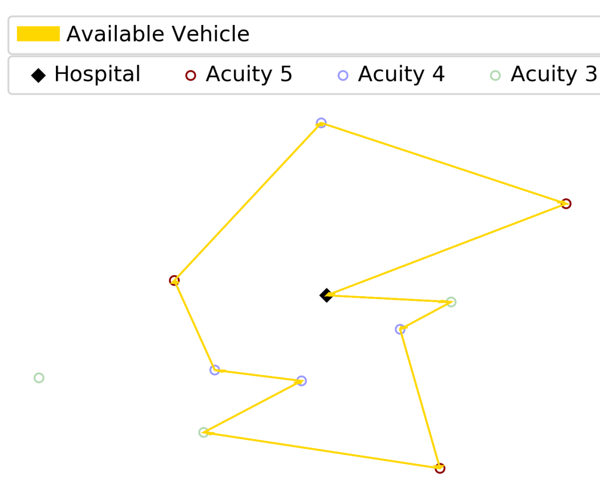
(d) Five hour time budget.



(e) Six hour time budget.



(f) Seven hour time budget.



time necessary, is larger. If two vehicles are available, the time required for supplemental vehicles is reduced when compared to the same run with only one available vehicle. For the runs depicted in Figure 2, if there are two available vehicles, and each with at least a nine hour time budget, no supplemental resources are necessary for visiting emergent patients with acuity level five.

These findings reveal the following insight: our model can identify which patients, including those that are emergent, are unable to be visited with available resources. This allows the hospital system to either allocate additional resources to visit such patients, as well as to contact such patients (via phone or email) to check up on their health condition to gauge whether they can be visited at a later time period.

Optimality Analysis. We assess computational times and optimality gaps while varying key parameters affecting the problem size, to demonstrate the computational performance of our optimization model. Figure 3 is divided into two sections by the dashed vertical line. For instances in Table 2, the left section indicates the computational run time in hours, while the right section depicts the final optimality gap achieved after 10 hours of run time. The green line corresponds to smaller instances with 10 to 50 patients, whereas the blue line represents instances with 60 to 100 patients. The y-axis is the number of instances either solved by the corresponding time (left section), or with an optimality gap less than or equal to the values on the x-axis (right section). As observed in Figure 3, the optimality gap is no more than 15% in the worst case. Among instances with 10 to 50 patients, the large majority of them ($\approx$ 81%) are solved to global optimality via Gurobi in less than 2 hours, and around 87% are solved to global optimality within 10 hour time limit with optimality gap $\approx$ 4% in the worst case. In instances with 60 to 100 patients, the average optimality gap over all instances is $\approx$ 3%, and it is $\approx$ 4% over instances that are not solved to global optimality within a 10 hour time limit.

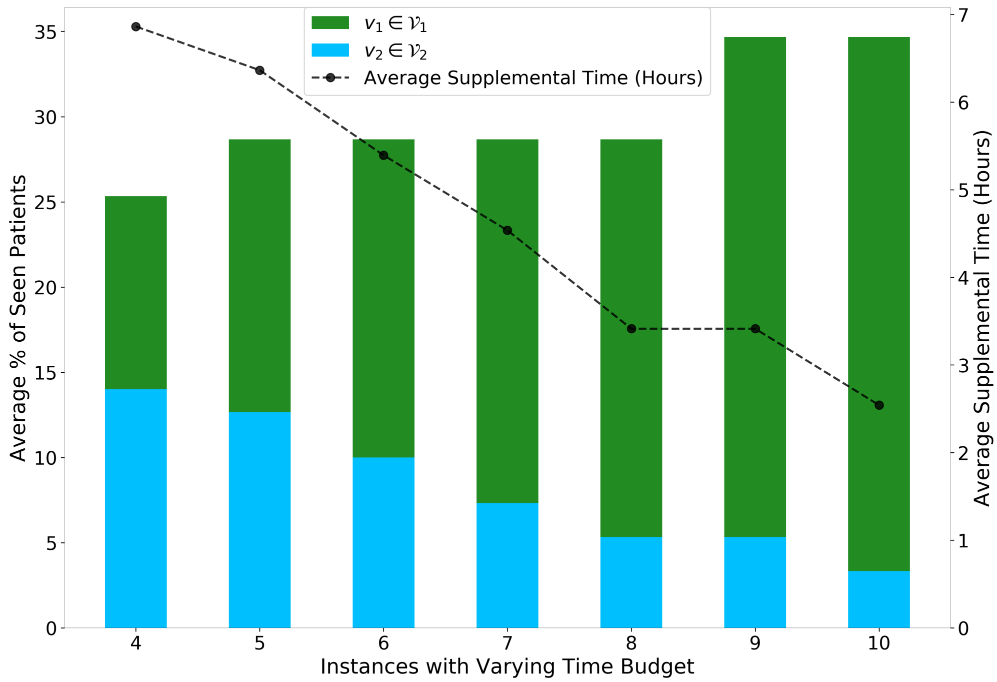
We now separate the results of larger instances with 60, 70, 80, 90, and 100 patients for two , three and four vehicles in Figure 4. It appears that the solution time is highly sensitive to the number of vehicles. While a majority of instances ($\approx$ 74%) with two vehicles are solved to global optimality within the time limit, the same can be said for only several instances with three and four vehicles. Over all instances with three and four vehicles that are not solved to global optimality within the time limit, the average optimality gap is 3.9%. As can be clearly seen in Figure 4, a vertical jump occurs to the right of the 0% optimality gap and adjacent to it. This suggest that nearly all of the gain in the first objective component of (weighted) patients has been achieved, and only minor distance improvements are being fine-tuned. We also investigate the performance of the model for various levels of patients in Appendix B.

These findings reveal the following insight: the considered instances are able to be solved, in reasonable times, to (near-)optimality, demonstrating the practicality of using mixed-integer optimization to inform managerial decision making concerning key features of CP programs.

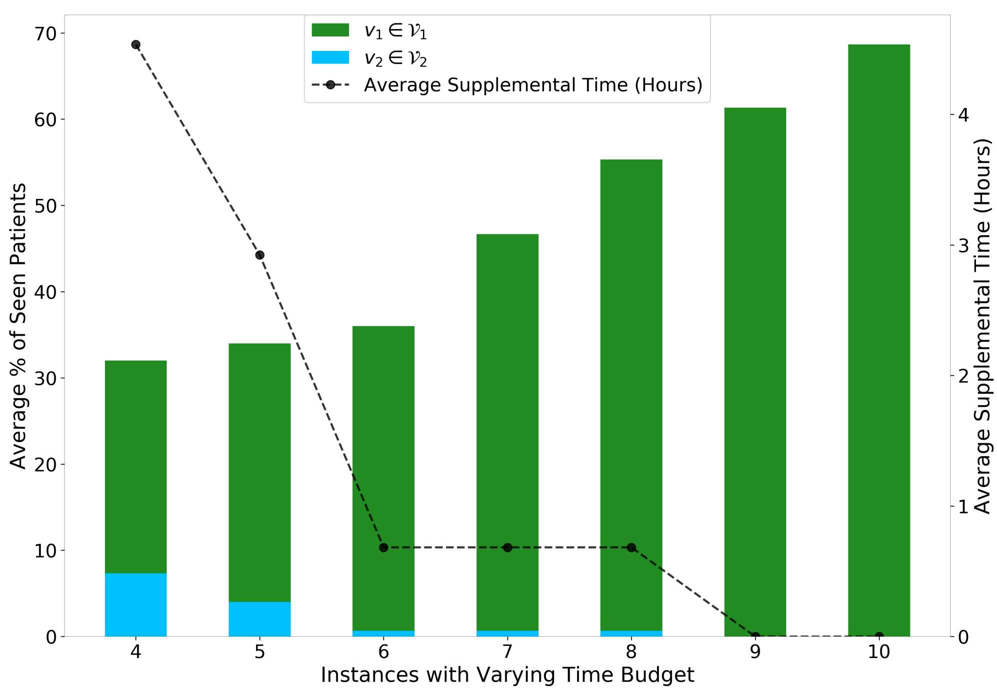
Experiments with Patient Time Windows. We now generate synthetic patient time windows to investigate the performance of our optimization model under the scenario that patients may specify their preferred time for a visit by a community paramedic. We consider an eight-hour working day beginning at 9:00 and divide it into two time windows: [9:00, 13:00] and [13:00, 17:00]. We consider three conditions: (1) In the first with *"None"* time windows, we set the patient availability time windows to be at their widest settings from 9:00 to 17:00; (2) In the second with *"Mild"* time windows, we randomly allocate these two time windows to half of the patients, and we assume that the remaining patients are available for visits during the entire eight-hour working day; (3) In the third with *"Strict"* time windows, we assign a time window to each patient randomly, reflecting the scenario where all patients have preferred time window for visitation.

Emergent patient preferences for a specific time window may be high in demand. Even so, the capability of our model to determine necessary supplemental resources guarantees feasibility. Thus,

Figure 2: Analysis of Available and Supplemental Resources for Instances with 50 Patients.



(a) One Available Vehicle and One Supplemental Vehicle.



(b) Two Available Vehicles and One Supplemental Vehicle.

Figure 3: Cumulative distribution plot of Gurobi performance on optimization formulation (4) for various instances detailed in Table 2. The green line depicts the results for smaller instances and the blue line demonstrates the performance of model for larger instances. Note that total number of smaller instances is 177, and total number of larger instances is 315.

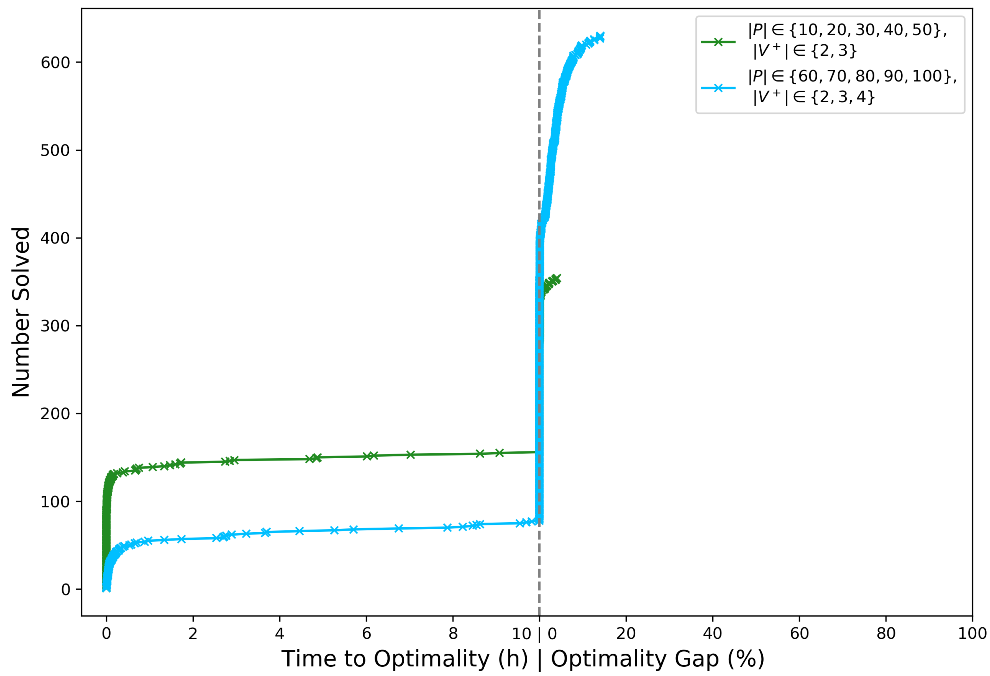
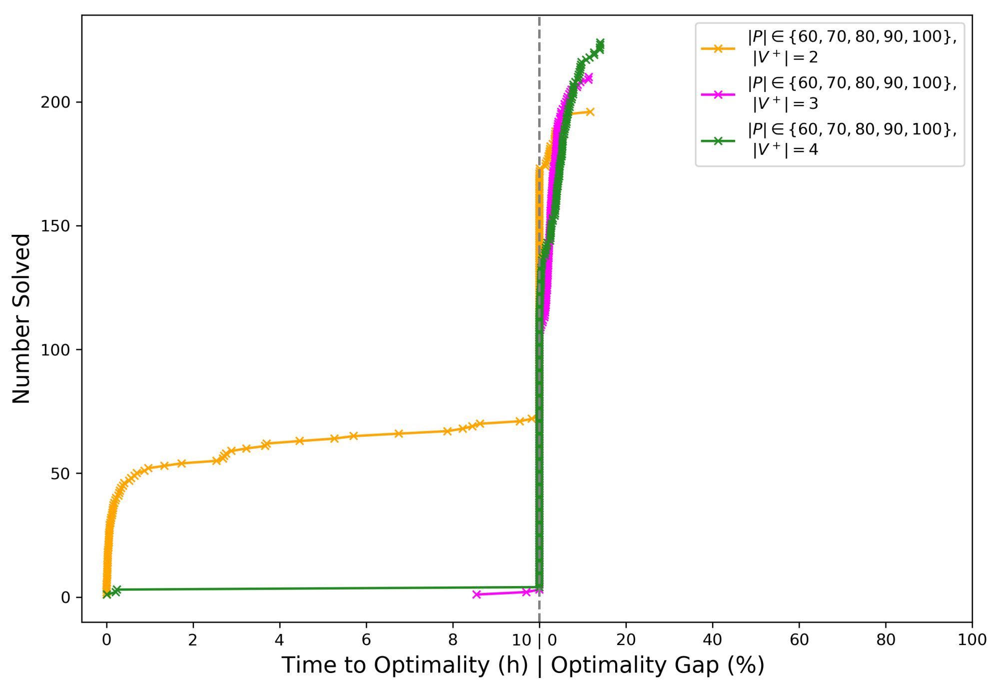


Figure 4: Cumulative distribution plot of Gurobi performance on optimization formulation (4) for instances with 60 to 100 patients and varying number of vehicles detailed in Table 2. The orange, purple, and green lines demonstrate the performance of model for instances with two, three, and four vehicles, respectively.



in this study we allow for multiple supplemental vehicles $v \in \mathcal{V}_2$, up to two supplemental vehicles when necessary. We assume that $|\mathcal{P}| \in \{50, 100\}$ and solve our optimization formulation (4) for the instances with these parameters. The optimality gap is below 10% for a full 93% of the experiments, with even the worst case instance having an optimality gap of less than 17.3%. Table 5 details summary results for instances with patient discharge times in June.

These findings reveal the following insight: time window restrictions affect the percentage of seen patients via available and supplemental vehicles. The percentage of seen patients via supplemental vehicles may increase in scenarios with stricter time windows because it is not possible to visit some of the emergent patients who were visited via available vehicles in scenarios with less strict time windows. These emergent patients are thus visited via supplemental vehicles, resulting in larger unused time budgets for available vehicles. This additional time is used to visit more patients with lower acuity levels. When the number of patients is fixed, as the number of vehicles increases, the percentage of seen patients via available vehicles increases, the percentage of seen patients via supplemental vehicles decreases, and the optimality gap often increases.

Table 5: Computational results for experiments with patient time windows for varying number of patients and vehicles, while time budget for each available vehicle is eight hours

; a "–" in *"Runtime"* column indicates reaching time limit prior to solving to optimality.

# Patients	# Available Vehicle(s)	Time Window	Time used Supplemental Vehicles		Seen Patients via Vehicle(s)		Runtime (Seconds)	Time to Find Best Incumbent (Seconds)	Optimality Gap
			Vehicle 1 (Hours)	Vehicle 2 (Hours)	Available	Supplemental			
50	1	None	4.14	0	11 (22.0%)	3 (6.0%)	162	120	0%
50	1	Mild	4.72	0	12 (24.0%)	4 (8.0%)	173.73	77	0%
50	1	Strict	4.32	0	10 (20.0%)	3 (6.0%)	2.27	0	0%
50	2	None	2.05	0	27 (54.0%)	1 (2.0%)	–	34,442	3.73%
50	2	Mild	2.05	0	25 (50.0%)	1 (2.0%)	–	30,904	7.60%
50	2	Strict	2.05	0	25 (50.0%)	1 (2.0%)	17,958	2,186	0.01%
100	1	None	12.89	0	12 (12.0%)	14 (14.0%)	–	19,629	5.07%
100	1	Mild	14.52	0	12 (12.0%)	15 (15.0%)	–	31,831	10.19%
100	1	Strict	9.10	6.35	13 (13.0%)	16 (16.0%)	–	32,624	3.58%
100	2	None	7.09	0	25 (25.0%)	6 (6.0%)	–	13,031	6.18%
100	2	Mild	5.73	3.97	27 (27.0%)	9 (9.0%)	–	33,482	6.79%
100	2	Strict	5.73	4.31	26 (26.0%)	9 (9.0%)	–	9,114	8.28%
100	3	None	5.16	0	41 (41.0%)	4 (4.0%)	–	16,079	9.75%
100	3	Mild	7.36	0	41 (41.0%)	6 (6.0%)	–	33,685	17.21%
100	3	Strict	6.53	0	39 (39.0%)	5 (5.0%)	–	30,083	8.30%

Experiments with Stochastic Visit Duration. We now explore the effect of stochastic visit durations by running experiments with visit durations that follow truncated lognormal distributions. Hoot et al. [34] considered the lognormal distribution for modeling ED patient evaluation and treatment times. As CP patients also require evaluation and treatment procedures for their medical needs, in our experiments we elect to model uncertain service times with the truncated lognormal distribution. We determine appropriate mean values for the truncated lognormal distribution that, after truncation, result in mean values that approach their deterministic counterparts, namely acuity levels 3, 4, and 5 have means approximately equal to 15, 20, and 30 minutes using truncation ranges of [10, 100], [15, 120], and [20, 140], respectively.

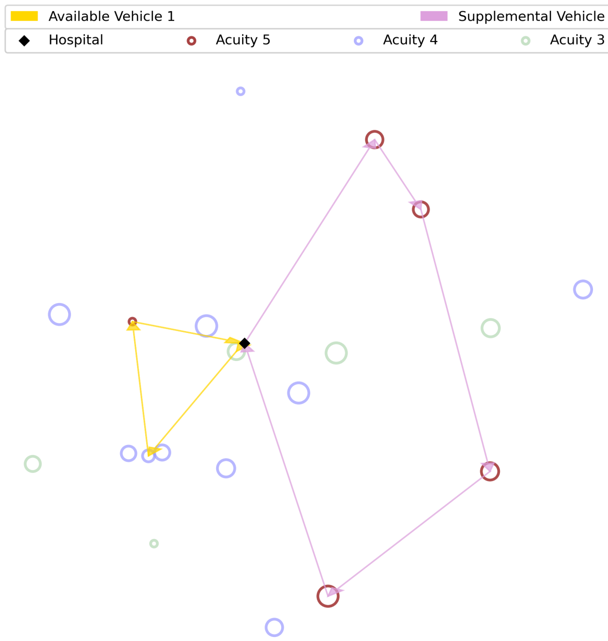
Table 6 shows the mean visit time of patients visited by available resources, supplemental resources, and unseen patients, per acuity level. Means are calculated over instances where there were positive values. This table reveals the following insights: emergent patients with longer duration times tend to be visited by supplemental resources. Moreover, there is a tendency for non-urgent patients to be unseen as expected visitation duration increases. Figure 5 reveals optimal routes of vehicles under stochastic visit durations between patients. Larger diameter nodes indicate longer visit durations. Figure 5 illustrates another useful insight: supplemental vehicles are used to visit nodes that have longer expected service durations, are more distant from other patient routes, or both.

Table 6: Computational Results for Experiments with Stochastic Visit Duration; Mean Duration Times (Min).

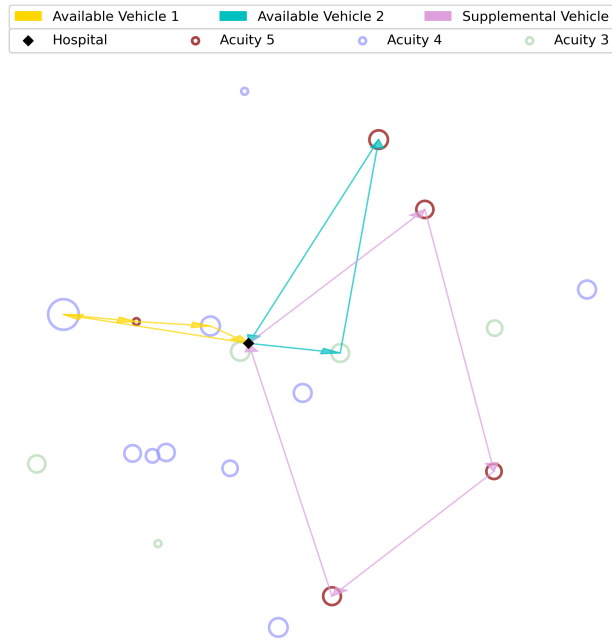
# Patients	# Vehicles	Time Budget	Mean Visit Duration of Seen Patients via Available Resources (Min)			Mean Visit Duration of Seen Patients via Supplemental Resources (Min)	Mean Visit Duration of Unseen Patients (Min)		Supplemental Vehicle Time (Hours)
			Acuity 5	Acuity 4	Acuity 3	Acuity 5	Acuity 4	Acuity 3	
10	1	2	23.90	27.27	-	52.38	30.72	22.85	2.47
10	2	2	27.76	18.78	17.32	68.60	26.40	19.74	1.81
20	1	2	28.18	20.59	10.96	50.56	32.26	24.82	1.60
20	1	3	29.38	24.47	17.27	33.27	23.33	25.81	2.27
20	1	4	28.43	18.11	12.05	60.80	34.71	32.73	1.27
20	2	2	26.91	21.93	10.09	89.18	27.54	23.25	1.89
20	2	3	32.56	20.00	13.40	48.63	33.51	18.65	2.12
20	2	4	43.97	22.18	16.29	107.15	39.02	26.08	2.24
50	1	8	32.18	18.54	-	58.83	28.39	21.55	6.92
50	2	8	41.90	21.73	11.17	-	35.90	20.40	-

Figure 5: Comparing optimal routes of supplemental vehicle for illustrative 20 patient instances.

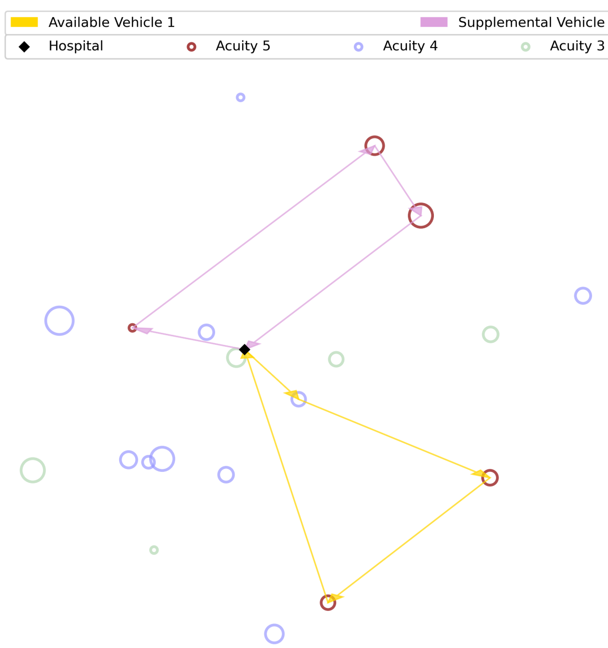
(a) Two hour time budget and one vehicle.



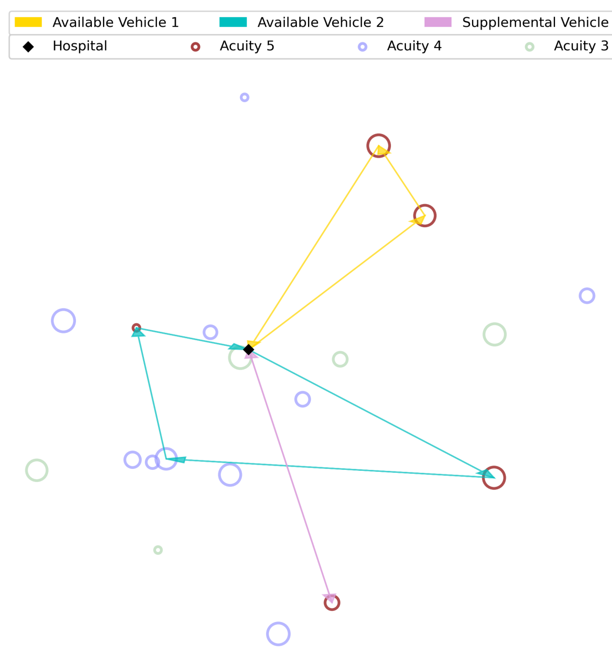
(b) Two hour time budget and two vehicles.



(c) Three hour time budget and one vehicle.



(d) Three hour time budget and two vehicles.



5. Concluding Remarks

Community paramedicine is a recent healthcare innovation that proactively treats chronic disease and potential ED repeat visits and inpatient readmissions at home, thereby preventing unnecessary ED visits and unplanned hospital readmissions. To date, only manual selecting, routing and scheduling approaches exist to tackle CP delivery challenges. This study establishes the first optimization-based framework to investigate the management of efficiency within a community paramedicine program. Through a broad set of computational experiments we validate the ability of our model to derive managerial insights for a variety of key questions related to the feasibility, efficiency and viability of community paramedicine programs.

We elaborate on several key insights. First, we highlight the ability of our modeling to reveal the needed amount of available resources to visit a specific percentage of patients, from 100% or less if desired or dictated by resource limitations. Second, while only emergent patients are visited when resources levels are low, less severe patients are visited when available resources are relatively high. Third, we highlight that our model can identify patients that are not visited via available resources, so as to either use supplemental resources to visit a patient (if emergent), or to contact and evaluate whether such patients can be visited at a later time period. Fourth, our mixed-integer optimization formulation (4) finds (near-)optimal tours in reasonable times, signifying its ability in a rural CP practice setting to answer key managerial questions. Fifth, with stricter patient time windows, supplemental vehicles may be responsible to visit an increased percentage of emergent patients; also, the percentage of seen patients with lower acuity levels by available vehicles may increase. Sixth, patients who are emergent and have longer expected visit duration, are more distant from other patient routes, or both, are more likely to be visited by supplemental resources, increasing their utilization. Seventh, patients not in urgent need of care are less likely to be seen by caregivers as the expected visit duration increases.

Figure 1 substantiates our results: for every vehicle, we clearly demonstrate the optimal selection and order for visiting priority-weighted patients, while favoring shorter tours. Moreover, we ensure that all mandatory patients are visited by existing resources, or if not, then we provide a solution with the minimal supplemental resources needed to do so. We propose a new objective function to improve societal health by maximizing the collected benefit of visiting patients, while secondarily favoring shorter tours for healthcare providers. The benefit is related to patient health features, and we propose a novel functional representation that preserves the priority of visiting more emergent patients.

We also address the notion of fairness in selecting patients, by prioritizing patients who are emergent according to our customized representation. Our approach has important societal implications for ensuring the overall welfare of our communities, and also provides for the treatment of a class of emergent patients who must be visited over the planning period. Moreover, our approach can accommodate for planning periods of a single day or potentially longer, in the following manner: the number of vehicles is multiplied by the number of days in the considered planning period, with the product equal to the number of available vehicles. To accommodate the situation when available resources are insufficient for visiting emergent patients, we propose to use supplemental vehicles. The supplemental vehicles guarantee model feasibility and serves a dual purpose – finding the minimal supplemental resources required to visit emergent patients and must be visited during the planning period. As CP is a relatively new innovation and the quantity of resources required for implementing a CP program is often uncertain, our model can help emergency medical service and hospital management determine the required resources for starting a CP program and, as appropriate, expand and improve an existing program by adding supplemental resources.

There are some limitations to the present study. Hospital systems, especially rural ones, are budget-strapped and may not have resources available to purchase state-of-the-art solvers. It remains an open question whether open source technologies can perform reasonably well on such problems. While our study focuses on the optimization aspect of CP programs, we acknowledge that establishing the feasibility of such a program is a critical precursor. This study assumes the existence of a functional CP program, and we recognize the importance of considering practical aspects such as personnel availability, training, and budget constraints when implementing a CP program. Our model has been developed within a specific context and its customization to different settings remains an open question. Factors such as local healthcare policies, patient demographics, and the availability of resources might necessitate adjustments to the model to ensure its applicability in other contexts. Our study considers a limited number of patients within a certain geographical region. Handling a larger number of patients, especially in smaller geographical areas, presents another open question. Additionally, our model may need to be adapted to account for different forms of transportation, such as public transit or walking, which could impact travel times and resource allocation.

More computational testing and algorithmic development would be needed to scale up to larger healthcare settings. In practice, such problems of the order presented in Table 5 can be solved overnight, which is within normal operating timeframes in the decision-making process (with supplemental vehicles serving to avoid infeasibility). As a result, we feel that our approach remains practically applicable for real-life scenarios. While those are the largest addressed in our manuscript, there are a few strategies that can be considered to manage the computational time for even larger instances. One possibility is to divide the problem into smaller, more manageable parts. This can be done by separating the problem based on different drivers, patients, or geographical areas, solve each partition individually, and then reconstruct an overall solution. Another way to handle large problems would be to modify the optimality gap according to the decision-maker's preferences. By allowing a slightly larger optimality gap, it is possible to reduce the computational time needed to obtain a near-optimal solution, while still providing valuable insights for decision-making.

Community paramedicine belongs to the Emergency Medical Service Agenda 2050 [14]. Using analytical approaches to design and implement CP programs is critical for efficient delivery of the future version of emergency medical service in the United States, and it opens a number of directions for future work. The integration of this idea into a rolling horizon to deal with uncertainty associated patient priority score and travel time between nodes, while intriguing, remains in the scope of future work. Our case study assumes the existence of one or more dedicated vehicles for CP. Moreover, should the same personnel or vehicles provide a variety of emergency medical services, such as responding to unpredictable true emergencies and doing CP visits, this would warrant additional investigation.

Appendix

A. Finding The Minimum Number of Supplemental Vehicles

We present Algorithm A1 to find the minimum number of supplemental vehicle(s) to visit all emergent patients who are not visited via available vehicles. Let δ represents the minimum number of needed supplemental vehicles. We use instances from Table 4 where the minimum time used by the supplemental vehicle is more than eight hours, to test the algorithm. We also compare the results for two approaches: i) We allow the model to use multiple supplemental vehicles with an eight hour time budget and find the minimum number of needed supplemental vehicles δ via Algorithm A1; ii) We assume there exists a single supplemental vehicle with a 24 hour time budget. In the former experiment, we assume all available and supplemental vehicles have an eight hour time budget.

The results in Table A.1 demonstrate that should the minimum extra time needed to visit all emergent patients turn out to be approximately eight hours in approach ii), the minimum number of supplemental vehicles δ is equal to the results of approach i). The reason for this behavior is that the total percentage of seen patients reduces in the results of approach i), as some of the patients with a lower acuity level are not visited via available vehicle(s) as compared to approach ii). Indeed, the model is more conservative in using supplemental vehicle(s) because the value of β is larger in approach i).

In the instances for which the total supplemental time is much larger than eight hours (e.g. row five in Table A.1), and the minimum number of supplemental vehicles is equal to two vehicles in approach i), the total time used by two vehicles is a bit larger as compared to approach i). Generally, the special structure of the model finds the minimum extra needed resources, and from that information the minimum number of supplemental vehicle(s) can be easily estimated.

Algorithm A1 Determine Minimum Number of Supplemental Vehicle(s)

```
Input : δ = 0
do
    Status = Solve optimization formulation (4a)–(4l) (δ)
    if Status = Infeasible then
        δ = δ + 1
while Status = Infeasible;
```

Table A.1: Minimum Number of Supplemental Vehicle(s) via Algorithm A1, versus One Supplemental Vehicle with Large Time Budget.

# Patients	# Available Vehicle(s)	Algorithm A1?	# Supplemental Vehicle(s)	Time Used Supplemental Vehicles		Seen Patients via Vehicle(s)		Runtime (Seconds)	Time to Find Best Incumbent (Seconds)	Optimality Gap
				Vehicle 1 (Hours)	Vehicle 2 (Hours)	Available	Supplemental			
80	1	Yes	1	7.98	-	15.0%	12.5%	-	2,745	2.12%
		No	1	9.06	-	17.5%	13.8%	1,380	1,261	0.00%
90	1	Yes	2	7.71	5.39	14.4%	16.7%	6,393	4,800	0.00%
		No	1	12.52	-	14.4%	16.7%	3,472	1,411	0.00%
100	1	Yes	2	7.78	7.93	13.0%	19.0%	34,169	13,353	0.01%
		No	1	15.15	-	13.0%	19.0%	9,965	9,538	0.01%
100	2	Yes	1	7.64	-	24.0%	8.0%	-	27,002	3.08%
		No	1	8.46	-	26.0%	9.0%	-	31,087	2.36%

B. Computational Results

In this section, we show the experimental results of solving formulation (4) for instances with 20, 40, 60, 80, and 100 patients, varying number of vehicles and time budget values that are even. In the following tables, each instance is named according to the convention *"Instance group– Number of patients – Number of vehicles – Time budget for each available vehicle"* to show the considered levels of varying parameters. Instance group is one, two, or three referring to discharge time of patients were in January, June, or October, respectively. In the following tables, a "–" in "Runtime" column indicates reaching time limit prior to solving to optimality.

Table B.2: Results for 20 Patients, Varying Number of Vehicles and Time Budgets.

Problem Code	Supplemental Time (Hours)	Seen Patients via Available Vehicle(s)	Seen Patients via Supplemental Vehicle	Runtime (Seconds)	Time to Find Best Incumbent (Seconds)	Optimality Gap
1–20–1–2	2.43	15.0%	15.0%	0.8	0	0.00%
1–20–1–4	0	25.0%	0%	1.4	0	0.00%
1–20–1–6	0	45.0%	0%	1.0	0	0.01%
1–20–1–8	0	65.0%	0%	2.6	2	0.00%
1–20–2–2	1.10	30.0%	5.0%	6.6	2	0.00%
1–20–2–4	0	55.0%	0%	3.7	3	0.00%
1–20–2–6	0	100.0%	0%	30.9	30	0.01%
2–20–1–2	4.39	15.0%	20.0%	1.5	1	0.00%
2–20–1–4	2.43	25.0%	10.0%	2.1	2	0.00%
2–20–1–6	0	35.0%	0%	1.2	1	0.00%
2–20–1–8	0	55.0%	0%	2.8	1	0.00%
2–20–2–2	3.87	25.0%	15.0%	3.2	1	0.00%
2–20–2–4	0	55.0%	0%	25.2	20	0.00%
2–20–2–6	0	90.0%	0%	12.0	11	0.01%
3–20–1–2	0	10.0%	0%	0.9	0	0.00%
3–20–1–4	0	30.0%	0%	1.0	0	0.00%
3–20–1–6	0	45.0%	0%	3.1	1	0.00%
3–20–1–8	0	65.0%	0%	50.6	32	0.00%
3–20–2–2	0	30.0%	0%	6.7	4	0.00%
3–20–2–4	0	60.0%	0%	29.5	29	0.01%
3–20–2–6	0	95.0%	0%	2,529	565	0.01%

Table B.3: Results for 40 Patients, Varying Number of Vehicles and Time Budgets.

Problem Code	Supplemental Time (Hours)	Seen Patients via Available Vehicle(s)	Seen Patients via Supplemental Vehicle	Runtime (Seconds)	Time to Find Best Incumbent (Seconds)	Optimality Gap
1–40–1–4	5.18	17.5%	15.0%	4.0	4	0.00%
1–40–1–6	3.51	22.5%	7.5%	4.4	3	0.00%
1–40–1–8	2.40	35.0%	5.0%	5.1	5	0.00%
1–40–1–10	1.69	45.0%	2.5%	6.7	6	0.00%
1–40–2–4	2.79	32.5%	5.0%	235.5	161	0.00%
1–40–2–6	0	50.0%	0%	2,404	1,053	0.00%
1–40–2–8	0	65.0%	0%	5,730	1,827	0.01%
1–40–2–10	0	87.5%	0%	–	8,547	0.25%
2–40–1–4	5.63	12.5%	12.5%	4.0	3	0.00%
2–40–1–6	4.14	20.0%	7.5%	5.4	4	0.00%
2–40–1–8	2.05	25.0%	2.5%	56.4	55	0.00%
2–40–1–10	2.05	40.0%	2.5%	296.1	121	0.00%
2–40–2–4	2.05	20.0%	2.5%	388.9	119	0.00%
2–40–2–6	2.05	47.5%	2.5%	–	29,753	3.16%
2–40–2–8	0	62.5%	0%	–	6,729	0.44%
2–40–2–10	0	87.5%	0%	–	8,356	0.18%
3–40–1–4	4.26	12.5%	7.5%	2.0	1	0.00%
3–40–1–6	2.54	22.5%	5.0%	5.0	4	0.00%
3–40–1–8	3.72	35.0%	5.0%	160.8	160	0.00%
3–40–1–10	1.77	42.5%	2.5%	478.8	234	0.00%
3–40–2–4	3.72	32.5%	5.0%	398.2	391	0.00%
3–40–2–6	0	47.5%	0%	4,808	740	0.01%
3–40–2–8	0	65.0%	0%	32,675	13,473	0.01%
3–40–2–10	0	92.5%	0%	–	21,417	0.17%

Table B.4: Results for 60 Patients, Varying Number of Vehicles and Time Budgets.

Problem Code	Supplemental Time (Hours)	Seen Patients via Available Vehicle(s)	Seen Patients via Supplemental Vehicle	Runtime (Seconds)	Time to Find Best Incumbent (Seconds)	Optimality Gap
1–60–1–6	7.54	15.0%	15.0%	196.5	196	0.00%
1–60–1–8	5.70	20.0%	10.0%	21.4	18	0.00%
1–60–1–10	4.90	28.3%	8.3%	208.6	108	0.00%
1–60–1–12	2.77	33.3%	5.0%	222.6	82	0.00%
1–60–2–6	3.02	31.7%	5.0%	30,785	10,189	0.01%
1–60–2–8	0	43.3%	0%	–	34,913	1.49%
1–60–2–10	0	58.3%	0%	–	4,116	1.45%
1–60–2–12	0	70.0%	0%	–	14,766	1.37%
1–60–3–6	0	51.7%	0%	–	30,894	4.44%
1–60–3–8	0	66.7%	0%	–	26,821	3.09%
1–60–3–10	0	88.3%	0%	–	24,180	0.33%
1–60–3–12	0	100.0%	0%	44.9	44	0.00%
2–60–1–6	5.59	13.3%	8.3%	58.3	16	0.00%
2–60–1–8	4.11	18.3%	5.0%	734.3	311	0.00%
2–60–1–10	3.49	25.0%	3.3%	588.3	421	0.00%
2–60–1–12	2.04	31.7%	1.7%	–	23,751	1.39%
2–60–2–6	2.04	28.3%	1.7%	–	31,791	3.75%
2–60–2–8	2.04	45.0%	1.7%	–	20,660	4.08%
2–60–2–10	0	56.7%	0%	–	17,299	2.50%
2–60–2–12	0	68.3%	0%	–	1,360	0.38%
2–60–3–6	0	48.3%	0%	–	10,326	6.97%
2–60–3–8	0	66.7%	0%	–	32,281	0.43%
2–60–3–10	0	91.7%	0%	–	35,106	0.15%
2–60–3–12	0	95.0%	0%	–	135	0.12%
3–60–1–6	5.87	15.0%	8.3%	9.3	8	0.00%
3–60–1–8	4.65	20.0%	5.0%	74.6	72	0.00%
3–60–1–10	2.93	26.7%	3.3%	351.7	351	0.00%
3–60–1–12	1.77	31.7%	1.7%	4,818	3,295	0.01%
3–60–2–6	4.65	35.0%	5.0%	–	29,052	3.39%
3–60–2–8	0	45.0%	0%	–	33,908	2.14%
3–60–2–10	0	56.7%	0%	–	35,486	2.72%
3–60–2–12	0	70.0%	0%	–	30,016	0.22%
3–60–3–6	0	50.0%	0%	–	14,388	6.38%
3–60–3–8	0	63.3%	0%	–	33,636	0.59%
3–60–3–10	0	96.7%	0%	–	35,422	0.08%
3–60–3–12	0	98.3%	0%	–	703	0.07%

Table B.5: Results for 80 Patients, Varying Number of Vehicles and Time Budgets.

Problem Code	Supplemental Time (Hours)	Seen Patients via Available Vehicle(s)	Seen Patients via Supplemental Vehicle	Runtime (Seconds)	Time to Find Best Incumbent (Seconds)	Optimality Gap
1–80–1–6	10.19	12.5%	16.3%	121	120	0.00%
1–80–1–8	9.06	17.5%	13.8%	1,379	1,261	0.00%
1–80–1–10	6.12	20.0%	8.8%	3,176	600	0.00%
1–80–1–12	4.21	22.5%	5.0%	122	106	0.00%
1–80–2–6	6.14	25.0%	8.8%	–	750	4.00%
1–80–2–8	3.67	33.8%	3.8%	–	29,157	2.56%
1–80–2–10	0	42.5%	0%	–	27,355	3.24%
1–80–2–12	0	53.8%	0%	–	22,256	1.09%
1–80–3–6	2.08	36.3%	2.5%	–	18,085	6.75%
1–80–3–8	0	52.5%	0%	–	22,340	3.92%
1–80–3–10	0	66.3%	0%	–	28,006	2.49%
1–80–3–12	0	86.3%	0%	–	35,060	1.34%
2–80–1–6	10.87	11.3%	15.0%	34,381	7,808	0.01%
2–80–1–8	9.03	13.8%	11.3%	13,202	11,111	0.00%
2–80–1–10	8.24	20.0%	10.0%	35,371	34,549	0.01%
2–80–1–12	5.77	22.5%	6.3%	28,346	19,414	0.01%
2–80–2–6	5.82	21.3%	6.3%	–	715	7.14%
2–80–2–8	3.75	31.3%	3.8%	–	23,435	3.63%
2–80–2–10	2.75	40.0%	1.3%	–	33,323	5.41%
2–80–2–12	2.75	52.5%	1.3%	–	29,762	3.64%
2–80–3–6	2.75	31.3%	1.3%	–	32,383	9.59%
2–80–3–8	2.75	48.8%	1.3%	–	34,090	6.99%
2–80–3–10	0	61.3%	0%	–	21,314	1.74%
2–80–3–12	0	78.8%	0%	–	29,129	1.87%
3–80–1–6	10.29	11.3%	15.0%	73.3	73	0.00%
3–80–1–8	9.19	16.3%	12.5%	78.8	46	0.00%
3–80–1–10	7.14	18.8%	8.8%	1,897	1,063	0.00%
3–80–1–12	4.86	21.3%	5.0%	24,316	20,026	0.01%
3–80–2–6	6.54	23.8%	7.5%	–	989	3.82%
3–80–2–8	3.71	32.5%	2.5%	–	35,898	3.10%
3–80–2–10	1.77	43.8%	1.3%	–	4,895	1.60%
3–80–2–12	0	52.5%	0%	–	2,603	1.74%
3–80–3–6	2.51	36.3%	2.5%	–	30,712	7.17%
3–80–3–8	1.99	52.5%	1.3%	–	12,463	5.32%
3–80–3–10	0	62.5%	0%	–	29,084	4.73%
3–80–3–12	0	81.3%	0%	–	34,743	0.38%

Table B.6: Results for 100 Patients, Varying Number of Vehicles and Time Budgets.

Problem Code	Supplemental Time (Hours)	Seen Patients via Available Vehicle(s)	Seen Patients via Supplemental Vehicle	Runtime (Seconds)	Time to Find Best Incumbent (Seconds)	Optimality Gap
1–100–1–6	17.13	11.0%	22.0%	452.1	451	0.01%
1–100–1–8	15.15	13.0%	19.0%	9,965.1	9,538	0.01%
1–100–1–10	13.41	17.0%	16.0%	1,496.4	1,486	0.00%
1–100–1–12	11.58	20.0%	13.0%	9,153.7	6,713	0.01%
1–100–2–6	11.58	19.0%	13.0%	–	14,274	0.77%
1–100–2–8	8.46	26.0%	9.0%	–	31,087	2.36%
1–100–2–10	6.44	34.0%	6.0%	–	1,307	2.16%
1–100–2–12	3.05	40.0%	3.0%	–	14,803	3.57%
1–100–3–6	8.00	29.0%	8.0%	–	19,239	5.36%
1–100–3–8	3.63	40.0%	3.0%	–	30,020	3.61%
1–100–3–10	2.68	52.0%	2.0%	–	4,837	5.29%
1–100–3–12	0	61.0%	0%	–	34,342	2.83%
2–100–1–6	15.34	10.0%	18.0%	–	35,850	6.34%
2–100–1–8	12.89	12.0%	14.0%	–	19,629	5.07%
2–100–1–10	12.24	17.0%	13.0%	–	18,749	3.59%
2–100–1–12	9.18	18.0%	9.0%	–	27,492	5.16%
2–100–2–6	12.24	21.0%	13.0%	–	1,013	9.47%
2–100–2–8	7.09	25.0%	6.0%	–	13,031	6.18%
2–100–2–10	5.16	33.0%	4.0%	–	27,835	5.00%
2–100–2–12	4.16	39.0%	2.0%	–	27,378	5.85%
2–100–3–6	7.72	29.0%	7.0%	–	30,520	10.78%
2–100–3–8	5.16	41.0%	4.0%	–	16,079	9.75%
2–100–3–10	2.99	51.0%	1.0%	–	7,138	4.67%
2–100–3–12	0	61.0%	0%	–	24,733	2.57%
3–100–1–6	13.07	10.0%	17.0%	31,061	22,859	0.01%
3–100–1–8	11.35	12.0%	14.0%	–	19,022	2.51%
3–100–1–10	9.45	15.0%	11.0%	–	35,479	1.78%
3–100–1–12	7.50	18.0%	8.0%	29,623	17,222	0.01%
3–100–2–6	8.17	18.0%	9.0%	–	14,742	6.38%
3–100–2–8	5.48	24.0%	4.0%	–	4,338	4.15%
3–100–2–10	1.99	32.0%	1.0%	–	2,027	2.84%
3–100–2–12	0	42.0%	0%	–	26,620	1.90%
3–100–3–6	6.35	30.0%	5.0%	–	32,015	7.41%
3–100–3–8	2.51	42.0%	2.0%	–	29,400	7.46%
3–100–3–10	0	56.0%	0%	–	27,416	4.08%
3–100–3–12	0	64.0%	0%	–	19,648	5.69%

In the following plots, we depict the cumulative distribution plot for instances with 60, 70, 80, 90, and 100 patients while varying the time budget for available vehicles. We separate the plots for two vehicles, three vehicles, and four vehicles.

Figure B.1: Cumulative distribution plot of Gurobi performance on optimization formulation (4) for instances with 60 to 100 patients and two vehicles. The blue, orange, green, red, and purple lines demonstrate the performance of model for instances with 60, 70, 80, 90, and 100 patients, respectively.

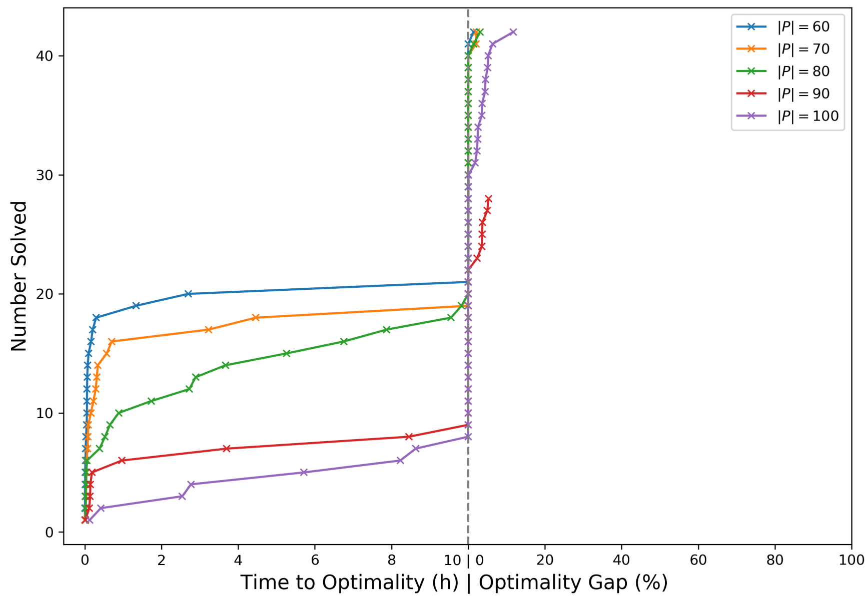


Figure B.2: Cumulative distribution plot of Gurobi performance on optimization formulation (4) for instances with 60 to 100 patients and three vehicles. The blue, orange, green, red, and purple lines demonstrate the performance of model for instances with 60, 70, 80, 90, and 100 patients, respectively.

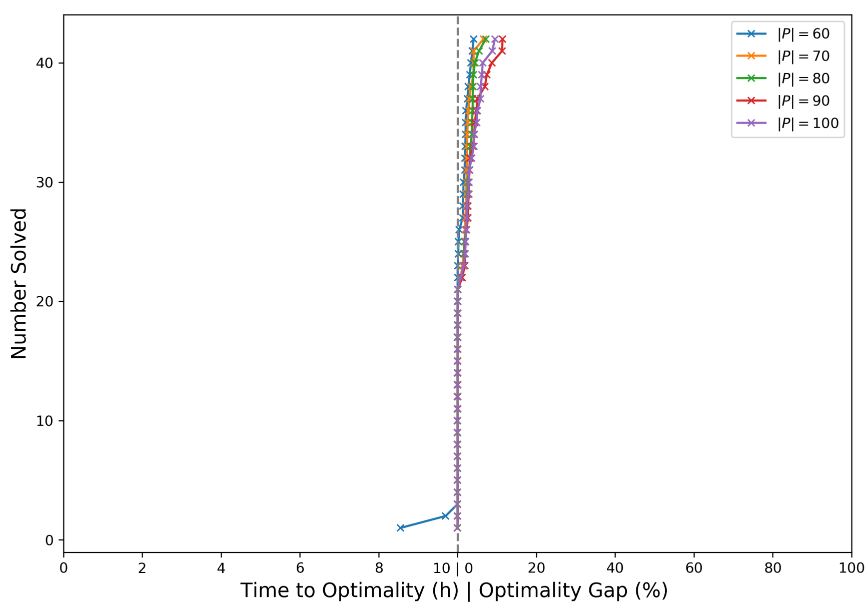
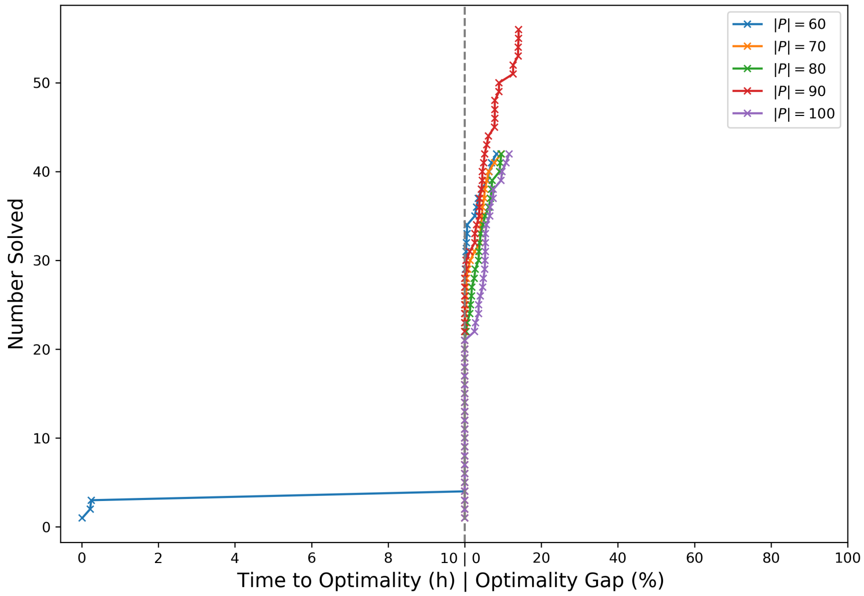


Figure B.3: Cumulative distribution plot of Gurobi performance on optimization formulation (4) for instances with 60 to 100 patients and four vehicles. The blue, orange, green, red, and purple lines demonstrate the performance of model for instances with 60, 70, 80, 90, and 100 patients, respectively.



C. Experiments with Longer Visit Durations.

We increase the visit duration for patients with acuity levels three, four, and five to 30, 60, 120 minutes, respectively. The results in Table C.7 show that the percentage of seen patients via available vehicles reduces remarkably, and the amount of supplemental vehicles significantly increases.

Table C.7: Results for Longer Visit Durations.

Problem Code	Supplemental Time (Hours)	Seen Patients via Available Vehicle(s)	Seen Patients via Supplemental Vehicle	Runtime (Seconds)	Optimality Gap
1-10-1-4	9.24	30.0%	40.0%	0.20	0%
1-10-1-6	6.93	30.0%	30.0%	0.21	0%
1-10-2-4	6.97	60.0%	30.0%	0.86	0%
1-20-1-4	9.24	15.0%	20.0%	0.47	0%
1-20-1-6	6.93	20.0%	15.0%	0.83	0%
1-20-1-8	4.90	20.0%	10.0%	0.46	0%
1-20-2-4	6.93	30.0%	15.0%	1.96	0%
1-20-2-6	2.60	35.0%	5.0%	2.33	0%
1-30-1-4	13.81	10.0%	20.0%	0.78	0%
1-30-1-6	11.89	13.3%	16.7%	1.58	0%
1-30-1-8	9.57	16.7%	13.3%	0.86	0%
1-30-1-10	7.45	16.7%	10.0%	1.23	0%
1-30-2-4	11.89	20.0%	16.7%	3.44	0%
1-30-2-6	7.45	23.3%	10.0%	12.81	0%
1-40-1-4	22.69	7.5%	25.0%	0.73	0%
1-40-1-6	20.91	10.0%	22.5%	1.58	0%
1-40-1-8	18.63	12.5%	20.0%	0.89	0%
1-40-1-10	16.28	12.5%	17.5%	6.79	0%
1-40-2-4	20.74	15.0%	22.5%	1.99	0%
1-40-2-6	16.78	20.0%	17.5%	1.31	0%
1-40-2-8	11.96	22.5%	12.5%	76.01	0%
1-40-2-10	7.52	25.0%	7.5%	61.99	0%

References

[1] Bill Fay. Emergency Rooms vs. Urgent Care Centers. https://www.debt.org/medical/emergency-room-urgent-care-costs/, 2019. Accessed: December 14, 2022.

[2] Eric Alper, Terrence A O'Malley, and Jeffrey Greenwald. Hospital discharge and readmission. https://www.uptodate.com/contents/hospital-discharge-and-readmission, 2020. Accessed: December 14, 2022.

[3] Jason H. Wasfy, Francesca Dominici, and Robert W. Yeh. An Obamacare success: Financial penalties reduce hospital readmission rates. https://www.statnews.com/2016/12/27/obamacare-success-penalties-readmissions/, 2016. Accessed: December 14, 2022.

[4] Kevin J Bennett, Matt W Yuen, and Melinda A Merrell. Community paramedicine applied in a rural community. *The Journal of Rural Health*, 34:s39–s47, 2018.

[5] Peter S Pang, Megan Litzau, Mark Liao, Jennifer Herron, Elizabeth Weinstein, Christopher Weaver, Dan O'Donnell, and Charles Miramonti. Limited data to support improved outcomes after community paramedicine intervention: A systematic review. *The American journal of emergency medicine*, 2019.

[6] American Academy of Orthopaedic Surgeons. *Community health paramedicine*, page 2. Jones & Bartlett Learning, 2017.

[7] International board of specialty certification. https://www.ibscertifications.org/about/about-us, 2009. Accessed: April 2023.

[8] American Academy of Orthopaedic Surgeons. *Community health paramedicine*, pages 3–4. Jones & Bartlett Learning, 2017.

[9] American Academy of Orthopaedic Surgeons. *Community health paramedicine*, pages 8–9. Jones & Bartlett Learning, 2017.

[10] Kent V Rondeau, Louis H Francescutti, and John J Zanardelli. Emergency department overcrowding: The impact of resource scarcity on physician job satisfaction/practitioner application. *Journal of Healthcare Management*, 50(5):327, 2005.

[11] Shakiba Enayati, Maria E Mayorga, Hector Toro-Díaz, and Laura A Albert. Identifying tradeoffs in equity and efficiency for simultaneously optimizing location and multipriority dispatch of ambulances. *International Transactions in Operational Research*, 26(2):415–438, 2019.

[12] Lisa I Iezzoni, Stephen C Dorner, and Toyin Ajayi. Community paramedicine—addressing questions as programs expand. *N Engl J Med*, 374(12):1107–9, 2016.

[13] Mattie Quinn. A New Kind of Paramedic for Less Urgent 911 Calls. https://www.governing.com/topics/health-human-services/gov-community-paramedicine-emergency-care.html, 2016. Accessed: December 14, 2022.

[14] EMS Agenda 2050. Ems Agenda 2050: Envisioning the Future. https://www.ems.gov/projects/ems-agenda-2050.html, 2019. Accessed: December 14, 2022.

[15] Karen Dewar. Livingston County Help for Seniors. https://www.ruralhealthinfo.org/project-examples/360, 2007. Date updated or reviewed: May 2, 2020. Accessed: December 14, 2022.

[16] Chase Wrangler. Rugby Community Paramedic Program. https://www.ruralhealthinfo.org/project-examples/988, 2017. Date updated or reviewed: December 14, 2018. Accessed: December 14, 2022.

[17] Sarah Calams. How 4 Community Paramedicine Programs are Positively Impacting Healthcare? https://bit.ly/3gnH7TN, 2018. Accessed: December 14, 2022.

[18] Paola Cappanera and Maria Grazia Scutellà. Joint assignment, scheduling, and routing models to home care optimization: A pattern-based approach. *Transportation Science*, 49(4):830–852, 2014.

[19] Paola Cappanera, Maria Grazia Scutellà, Federico Nervi, and Laura Galli. Demand uncertainty in robust home care optimization. *Omega*, 80:95–110, 2018.

[20] Matias Sevel Rasmussen, Tor Justesen, Anders Dohn, and Jesper Larsen. The home care crew scheduling problem: Preference-based visit clustering and temporal dependencies. *European Journal of Operational Research*, 219(3):598–610, 2012.

[21] Andrea Trautsamwieser, Manfred Gronalt, and Patrick Hirsch. Securing home health care in times of natural disasters. *OR spectrum*, 33(3):787–813, 2011.

[22] Salma Chahed, Eric Marcon, Evren Sahin, Dominique Feillet, and Yves Dallery. Exploring new operational research opportunities within the home care context: The chemotherapy at home. *Health Care Management Science*, 12(2):179–191, 2009.

[23] Patrik Eveborn, Patrik Flisberg, and Mikael Rönnqvist. LAPS care–an operational system for staff planning of home care. *European Journal of Operational Research*, 171(3):962–976, 2006.

[24] David Bredström and Mikael Rönnqvist. Combined vehicle routing and scheduling with temporal precedence and synchronization constraints. *European Journal of Operational Research*, 191(1):19–31, 2008.

[25] Stefan Bertels and Torsten Fahle. A hybrid setup for a hybrid scenario: Combining heuristics for the home health care problem. *Computers & Operations Research*, 33(10):2866–2890, 2006.

[26] Kristine Thomsen. Optimization on home care. Master's thesis, Technical University of Denmark, Denmark, 2006.

[27] Yoram Clapper, Joost Berkhout, René Bekker, and Dennis Moeke. A model-based evolutionary algorithm for home health care scheduling. *Computers & Operations Research*, page 106081, 2022.

[28] Erfaneh Nikzad, Mahdi Bashiri, and Babak Abbasi. A matheuristic algorithm for stochastic home health care planning. *European Journal of Operational Research*, 288(3):753–774, 2021.

[29] Maria Di Mascolo, Cléa Martinez, and Marie-Laure Espinouse. Routing and scheduling in home health care: A literature survey and bibliometric analysis. *Computers & Industrial Engineering*, 158:107255, 2021. ISSN 0360-8352. doi: https://doi.org/10.1016/j.cie.2021.107255.

[30] Biao Yuan, Ran Liu, and Zhibin Jiang. A branch-and-price algorithm for the home health care scheduling and routing problem with stochastic service times and skill requirements. *International Journal of Production Research*, 53(24):7450–7464, 2015.

[31] Yang Zhan, Zizhuo Wang, and Guohua Wan. Home service routing and appointment scheduling with stochastic service times. *European Journal of Operational Research*, 288(1):98–110, 2021.

[32] Yong Shi, Toufik Boudouh, Olivier Grunder, and Deyun Wang. Modeling and solving simultaneous delivery and pick-up problem with stochastic travel and service times in home health care. *Expert Systems with Applications*, 102:218–233, 2018.

[33] Ettore Lanzarone and Andrea Matta. A cost assignment policy for home care patients. *Flexible Services and Manufacturing Journal*, 24(4):465–495, 2012.

[34] Nathan R Hoot, Larry J LeBlanc, Ian Jones, Scott R Levin, Chuan Zhou, Cynthia S Gadd, and Dominik Aronsky. Forecasting emergency department crowding: a discrete event simulation. *Annals of emergency medicine*, 52(2):116–125, 2008.

[35] Christian Fikar and Patrick Hirsch. Home health care routing and scheduling: A review. *Computers & Operations Research*, 77:86–95, 2017.

[36] Gerhard Hiermann, Matthias Prandtstetter, Andrea Rendl, Jakob Puchinger, and Günther R Raidl. Metaheuristics for solving a multimodal home-healthcare scheduling problem. *Central European Journal of Operations Research*, 23(1):89–113, 2015.

[37] David Barrera, Nubia Velasco, and Ciro-Alberto Amaya. A network-based approach to the multi-activity combined timetabling and crew scheduling problem: Workforce scheduling for public health policy implementation. *Computers & Industrial Engineering*, 63(4):802–812, 2012.

[38] John Bowers, Helen Cheyne, Gillian Mould, and Miranda Page. Continuity of care in community midwifery. *Health care management science*, 18(2):195–204, 2015.

[39] Ahmet Cinar, F Sibel Salman, and Burcin Bozkaya. Prioritized single nurse routing and scheduling for home healthcare services. *European Journal of Operational Research*, 289(3): 867–878, 2021.

[40] Nastaran Oladzad-Abbasabady and Reza Tavakkoli-Moghaddam. Dynamic routing-scheduling problem for home health care considering caregiver-patient compatibility. *Computers & Operations Research*, 148:106000, 2022.

[41] Ahmet Cinar, F Sibel Salman, Ozgur M Araz, and Mert Parcaoglu. Managing home healthcare services with dynamic arrivals during a public health emergency. *IEEE Transactions on Engineering Management*, 2022.

[42] Faisal Alkaabneh and Ali Diabat. A multi-objective home healthcare delivery model and its solution using a branch-and-price algorithm and a two-stage meta-heuristic algorithm. *Transportation Research Part C: Emerging Technologies*, page 103838, 2022.

[43] Ran Liu, Xiaolan Xie, Vincent Augusto, and Carlos Rodriguez. Heuristic algorithms for a vehicle routing problem with simultaneous delivery and pickup and time windows in home health care. *European Journal of Operational Research*, 230(3):475–486, 2013.

[44] George B Dantzig and John H Ramser. The truck dispatching problem. *Management science*, 6(1):80–91, 1959.

[45] Pieter Vansteenwegen, Wouter Souffriau, and Dirk Van Oudheusden. The orienteering problem: A survey. *European Journal of Operational Research*, 209(1):1–10, 2011.

[46] Wouter Souffriau, Pieter Vansteenwegen, Greet Vanden Berghe, and Dirk Van Oudheusden. The multiconstraint team orienteering problem with multiple time windows. *Transportation Science*, 47(1):53–63, 2013.

[47] I-Ming Chao, Bruce L Golden, and Edward A Wasil. The team orienteering problem. *European journal of operational research*, 88(3):464–474, 1996.

[48] Gurobi. *Gurobi Optimizer 8.0.4 Reference Manual*. Gurobi Optimization, Inc., 2020.

[49] HERE Map API. https://www.here.com. Accessed: December 14, 2022.