

# On the Mass COVID-19 Vaccination Scheduling Problem

Chuang Zhang

Department of Equipment Support and Remanufacturing, Army Academy of Armored Forces, 100072 Beijing, China  
chuangzhang1225@163.com

Yantong Li\*

School of Maritime Economics and Management, Dalian Maritime University, Dalian 116026, China yantongli@163.com

Junhai Cao

Department of Equipment Support and Remanufacturing, Army Academy of Armored Forces, 100072 Beijing, China  
jhcao@163.com

Xin Wen

Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong  
windy.wen@polyu.edu.hk

The outbreak of COVID-19 dramatically impacts the global economy. Mass COVID-19 vaccination is widely regarded as the most promising way to fight against the pandemic and help return to normal. Many governments have authorized certain types of vaccines for mass vaccination by establishing appointment platforms. Mass vaccination poses a vital challenge to decision-makers responsible for scheduling a large number of appointments. This paper studies a vaccination site selection, appointment acceptance, appointment assignment, and scheduling problem for mass vaccination in response to COVID-19. An optimal solution to the problem determines the open vaccination sites, the set of accepted appointments, the assignment of accepted appointments to open vaccination sites, and the vaccination sequence at each site. The objective is to simultaneously minimize 1) the fixed cost for operating vaccination sites; 2) the traveling distance of vaccine recipients; 3) the appointment rejection cost; and 4) the vaccination tardiness cost. We formulate the problem as a mixed-integer linear program (MILP). Given the NP-hardness of the problem, we then develop an exact logic-based Benders decomposition (LBBD) method and a matheuristic method (MH) to solve practical-sized problem instances. We conduct numerical experiments on small- to large-sized instances to demonstrate the performance of the proposed model and solution methods. Computational results indicate that the proposed methods provide optimal solutions to small-sized instances and near-optimal solutions to large ones. In particular, the developed matheuristic can efficiently solve practical-sized instances with up to 500 appointments and 50 vaccination sites. We discuss managerial implications drawn from our results for the mass COVID-19 vaccination appointment scheduling, which help decision-makers make critical decisions.

*Key words:* COVID-19; Mass vaccination; Appointment scheduling; Logic-based Benders decomposition; Matheuristic.

---

\* Corresponding author

## 1. Introduction

The world has suffered from the Corona Virus Disease 2019 (COVID-19) since its outbreak in December 2019. More than 240 million infected cases have been reported, with nearly five million deaths confirmed (Worldometer). The COVID-19 pandemic has also resulted in vital global economic losses and significantly impacted social activities and livelihoods. Governments implement policies including massive testing, strict social distancing, and lockdowns to prevent the spread of the virus. In the high infectiousness and long incubation periods without apparent symptoms, it is imperative to implement vaccines, which the world is longing for. By November 2021, over 260 vaccines are in development, some of which have been approved by governments. (Milken Institute and FirstPerson) Many countries and regions have initiated mass vaccination programs and open channels through which residents can make vaccination appointments online. More than 220 million residents have been vaccinated, including 79.6% of the adults (Centers for Disease Control and Prevention). The UK has vaccinated 73% of its population (Our World in Data). However, data shows that till the beginning of October 2021, only 3.5% of people in low-income countries have received COVID-19 vaccine (Our World in Data). In those countries, the mass vaccination programs are still in the early stages due to the limited supply of the vaccine. Consequently, early vaccination will not be available for everyone.

The conflict between demand and capacity has caused some mess in the appointment scheduling program. Dai (2021) has pointed out that the traditional vaccine sign-up model performs poorly for COVID-19 vaccination appointment scheduling, as many people keep refreshing the websites hoping to get appointments. To coordinate the situation, governments choose to firstly deliver the vaccines to people with higher urgency and priority. The Arizona Department of Health Services has published the COVID-19 vaccine prioritization, which divides the vaccination program into three phases. The first phase, from December 2020 to spring of 2021, is allocated for people with high priorities, including healthcare workers & healthcare support occupations, emergency medical service workers, and people in long-term care facility staff & residents (Arizona Department of Health Services). Therefore, group appointments are made through appointment platforms by companies, organizations, or communities, which are geographically dispersed. When making the appointment, the individual (usually the administrator) who makes the appointment must specify the following information: 1) the number of personnel that needs vaccination; 2) the time window for vaccination during the day.

Decision-makers must make several critical decisions to establish a mass vaccination program, including which vaccination site to open, which vaccination appointment to accept or reject, how to assign the accepted vaccination appointments to opened vaccination sites, and the service time of each appointment at each site. These decisions are made under the consideration of several critical

factors. First, due to the limited budget, vaccination sites must be well launched and operated, such that a high resource utilization rate and a cost-effective plan can be achieved. Second, the storage and handling of COVID-19 vaccines have stringent requirements on temperature, venue, and other conditions. COVID-19 vaccines need to be stored or transported in cold chains and be placed in specific containers with a low temperature around  $2^{\circ}\text{C}$  to  $8^{\circ}\text{C}$ . Some even require temperature below  $-60^{\circ}\text{C}$  (Wang et al. 2020a). These requirements indicate that the number of facilities that can be selected as vaccination sites is limited. Third, decision-makers must provide vaccine recipients with good service, including a moderate travel distance to get vaccinated and an accurate service time that meets recipients' expected time window. Thus, the mass vaccination program consists of complex optimization problems with potentially conflicting objectives. For example, Suzhou, a big city with more than ten million residents in Jiangsu Province, China, has launched 179 COVID-19 vaccination sites, 44 of which are to be opened according to the situation (Suzhou Municipal People's Government). The operation of these vaccination sites occupies resources, e.g., venue, facilities, healthcare workers, and other staff. More vaccination sites lead to high operation costs and low resource utilization. At the same time, more vaccination sites indicate a high capacity level by providing vaccine recipients with easily accessible vaccination sites and short waiting times. On the contrary, decision-makers can save costs through opening fewer vaccination sites, which results in a long travel distance of vaccine recipients, a higher probability of appointment rejection, and service tardiness. Therefore, we are motivated by the above practical considerations and aims to provide complete and effective solutions with a tradeoff between the operational cost and the service level.

This paper investigates the mass COVID-19 vaccination appointment scheduling (MCVAS) problem, which jointly determines the vaccination site selection, vaccination appointment acceptance or rejection, vaccination appointment assignment to opened vaccination sites, and the service schedule at each vaccination site. The objective is to simultaneously minimize four components, i.e., the fixed cost for opening a vaccination site, total travel distance of vaccine recipients, total appointment rejection cost, and total tardiness cost. The research aims to provide a good balance between cost-effectiveness and service level. The focus is to generalize the classic appointment scheduling problem by considering multiple vaccination sites and a fixed cost for opening a vaccination site. It also extends the scheduling and location problem by considering appointment rejection. We formulate the problem as a mixed-integer linear programming (MILP) model adapted from similar scheduling problems. Due to the strong NP-hardness of the problem, we then develop an exact logic-based Benders decomposition (LBBD) method to solve it efficiently. The LBBD method decomposes the original problem into a master problem (MP) and a subproblem (SP). The MP deals with the vaccination site selection, appointment acceptance, and assignment decisions. Once a

feasible solution to the MP is obtained, the corresponding SP determines the vaccination sequences at each vaccination site. Benders cuts are iteratively generated and added to the MP. To tackle practical-sized instances with hundreds of appointments and dozens of vaccination sites, we propose a matheuristic method (MH) based on predetermined sequences of the appointments. Excessive computational experiments are conducted to evaluate the performances of the model and algorithms. Sensitivity analyses give quantitative insights into the effects of different cost components in the objective. Based on the computational results, we provide some managerial implications for the mass COVID-19 vaccination appointment scheduling decision-making.

The remainder of this paper is organized as follows. Section 2 overviews literature related to the proposed problem. Section 3 describes the problem in detail and presents the MILP model. The exact LBBD method and the matheuristic are developed in Section 4. Section 5 conducts the computational study, analyzes the results, and performs sensitivity analysis. Managerial implications are presented in Section 6. Section 7 summarizes the paper with conclusions.

## 2. Literature review

This paper is an attempt to apply operational research methods to promote mass vaccination scheduling in the COVID-19 pandemic. The studied problem is closely related to several research streams, including the vaccine supply chain management, appointment scheduling problem, parallel machine scheduling problem with rejection, and the scheduling and location (ScheLoc) problem. We next review these related topics and indicate how they differ from the current study.

### 2.1. Operational research methods applied in pandemics

Operational research (OR) methods play important roles in decision-making processes against pandemics, including severity assessment, demand forecasting, facility location, supply chain management, resource scheduling, among others. Bennett et al. (2012) outlined the roles that OR analysis played in the health policy in England. Among the three areas, the health protection issue focused on threats to public health, and the authors took influenza, the “Swine Flu” pandemic, and the infectious Creutzfeldt–Jakob Disease as examples to show that OR was critical in modeling and assessment. More potential contributions that OR could make were also reflected in the paper. Nagurney (2021) considered the challenges in labor shortage in supply chains due to the COVID-19 pandemic. A supply chain optimization model was proposed with a profit-maximization objective. The model defined labor as a critical variable and considered both fixed and elastic demand cases. Illustrative examples were presented to show the impacts of labor availability in supply chains. Zhao et al. (2021) focused on the infectious waste management in pandemics considering uncertain generation. A scenario-based bi-objective robust optimization model was formulated to simultaneously determine the location of waste management facilities and the routing of waste

collection and transportation. The objective functions contained both cost and risk minimization. Three multi-objective approaches were proposed for the problem, and a real-life case study in the COVID-19 pandemic was conducted. Several models are built to describe the spread of viruses and the effectiveness of various anti-epidemic measures. Among them, the SEIR model is one of the most important and widely used ones (Albani et al. 2021, Foy et al. 2021). The SEIR model divides people into four states, that is susceptible (S), exposed (E), infected (I), and recovered (R). The states of people may change dynamically with the spread of the epidemic, the measures of governments, and the treatments by medical workers. Jiao et al. (2020) presented an SEIR model for the transmission of COVID-19 pandemic and considered the infectivity of incubation period and homestead-isolation on the susceptible. The paper suggested the implementation of strict isolation to curb the propagation of the COVID-19 virus.

## **2.2. Vaccine supply chain management**

Vaccines are crucial to stop the spread of the virus, relieve social anxiety, and recover global economies. From the view of operations research, many researchers focus on the supply chain management concerning the vaccine transportation and distribution, the vaccination site location, and the prioritization of vaccine recipients. Hovav and Tsadikovich (2015) considered the inventory management and distribution of influenza vaccines in a multi-echelon healthcare supply chain. A mixed-integer programming (MIP) model considering recipient priorities was formulated to minimize the total costs. A case study of the vaccination program in Israel was presented to show the practical value of the paper. Lim et al. (2016) studied the vaccination outreach location selection problem in low and middle income countries. They proposed four coverage models to maximize the number of vaccinated residents. Li et al. (2020) studied the location-inventory problem for vaccination. A multi-objective mixed-integer nonlinear programming model was proposed to minimize the average travel distances, maximize the number of open vaccination stations, and minimize the total cost, simultaneously. The model was simplified to obtain a MILP using a two-stage strategy, solved by an  $\epsilon$ -constraint method. Albani et al. (2021) proposed a SEIR-like model to describe infection severity levels to evaluate the impacts of vaccination delay. They considered time-dependent parameters and control strategies to make the model accurate and realistic. The results showed that vaccination delay would severely affect mortality, hospitalization, and recovery projections. Georgiadis and Georgiadis (2021) built a novel framework for the optimal planning of both the vaccine supply chain management and the daily vaccine implementations in vaccination centers. A MILP model with cost minimization was proposed to determine the vaccine delivery, inventory control, and daily vaccination schedules. A decomposition-based approach was employed to solve large-scale problems, and a case study of Greek COVID-19 vaccine supply chain management was

conducted. The paper also designed a MILP-based replanning algorithm to deal with potential disturbances such as no show.

However, the literature on vaccine supply chain management rarely addresses the appointment scheduling problem, a critical optimization problem in vaccination programs.

### **2.3. Appointment scheduling**

Appointment scheduling and management is a common challenge in many industries, such as truck appointments at marine terminals (Huynh 2009), production systems (Elhafsi 2002), project scheduling (Bendavid and Golany 2011), and machine scheduling (Elmaghraby et al. 2005). The presented review of appointment scheduling focuses on healthcare industries to coordinate the conflict between limited medical resources and the seemingly endless demands.

The appointment scheduling studies in the medical field focus on single service provider settings, e.g., a healthcare service center. De Vuyst et al. (2014) studied the healthcare appointment scheduling problem where a single physician treated patients in a fixed-length session. The authors applied discrete-time setting and Lindley's recursion to design an efficient algorithm for schedule evaluation, in which patient waiting time and physician idle time were considered. Feldman et al. (2014) investigated an appointment scheduling problem considering patients' time preferences and a single service provider. The no-show behavior, indicating that patients with appointments fail to appear, was also considered in the problem. They presented static and dynamic programs to maximize the expected daily profit, which was the difference between revenue and service costs. The dynamic model considered the status of appointments and was solved by a policy improvement heuristic. Alizadeh Foroutan et al. (2020) studied a non-emergency outpatient appointment scheduling problem with a single machine and limited staff. A MILP model was proposed to minimize the penalties of patients' undesirable days, patients' undesirable hours, machine idle time, and doctor switches. A genetic algorithm was designed, which showed superiority to the MILP model through computational experiments.

The literature on the appointment scheduling problem with multiple service providers is limited. Zhou et al. (2020) focused on the optimization of patient-and-physician matching and appointment scheduling in specialty care to minimize the matching and operational costs. A two-stage formulation was proposed for the problem. The first stage was the patient-physician assignments, while the second one was about each service provider's classic appointment scheduling problems. The sample average approximation (SAA) method and an improved Benders decomposition method were designed for the problem. No-show consideration was also incorporated in the paper. Shnits et al. (2020) investigated appointment scheduling in the healthcare system with parallel servers and pre-sequenced patients. The probability of no-shows was considered. The objective is to minimize the end of the day and increase resource utilization under service quality requirements. A

stochastic formulation was proposed, which considers the probability of no-shows and stochastically distributed service time. A deterministic MILP model was first formulated to approximate the problem. Then, a sequential multi-server numerical-based algorithm was developed to overcome the dimensionality limitations of the MILP. Soltani et al. (2019) considered an appointment scheduling problem with multiple identical providers. The problem, taking stochastic service times and customer no-shows into consideration, was modeled as a time-inhomogeneous discrete-time Markov chain process to minimize the weighted sum of customers' waiting time, providers' idle time, and overtime. A load-based appointment scheduling heuristic was proposed to quickly find optimal or near-optimal schedules based on some optimal conditions. Machine learning techniques were employed to build the efficient and straightforward heuristic, learning from an extensive database of instances. The methods were evaluated through a real-world experiment at a local legal counseling center with 3919 appointments.

To sum up, the literature on appointment scheduling problems focuses on a single location, such as a hospital. There is a lack of flexibility for assigning personnel to different sites to reduce potential operational costs. However, in vaccination operations, it is essential to assign personnel to different vaccination sites to guarantee service quality, reduce operational costs, and avoid the potential risk of cross infections. Besides, classic appointment scheduling problems do not allow an appointment to be rejected. However, during the early stage of mass COVID-19 vaccination, the vast demand can hardly be met, and decision-makers must decide which appointment to reject based on prioritization rules. We next review the stream of research on parallel machine scheduling problems with rejection.

#### **2.4. Parallel machine scheduling with rejection**

The parallel machine scheduling (PMS) problem is well studied. In classic PMS problems, a set of jobs are assigned to multiple locations to be processed sequentially. The objectives that are commonly considered include the minimization of makespan (Li et al. 2011, Vallada and Ruiz 2011), total completion time (Bülbül and Şen 2016, Shim and Kim 2007), earliness and/or tardiness (Bektur and Sara 2019), and cost-related performance measures (Ji et al. 2013). Parallel machine scheduling with rejection was first studied in Bartal et al. (2000) where jobs might be rejected with a penalty cost in a classic PMS problem. Dósa and He (2006) investigated the problem where a set of jobs were to be processed while no machine was initially available. When a job popped up, the decision-makers needed to choose whether to reject it, process it on existing machines without preemption, or purchase a new machine to process it. The objective was to minimize makespan, machine purchasing costs, and rejection penalty costs. A two-phase online algorithm was proposed to solve the small job cases in which the size of a job was no more than the purchasing cost of a

machine. Ou et al. (2015) studied a classic PMS problem with job rejection to minimize the sum of makespan and rejection penalty. Several optimal properties were presented, and a heuristic with a worst-case bound was developed. Zhong and Ou (2017) studied the parallel machine scheduling problem with job rejection to minimize makespan and total penalty cost. A bound of maximum total processing time of the rejected jobs was predefined in the problem. Due to the strong NP-hardness of the problem, a 2-approximation heuristic algorithm was introduced based on some optimal properties, and a polynomial-time approximation scheme was proposed.

Most literature on PMS problems with job rejection assumes the machines are in the same location. In our problem, the number and locations of vaccination sites are neither fixed nor given but determined by decision-makers. However, we identify another research area related to our research which handles machine locations and scheduling simultaneously, denoted as the ScheLoc problem.

## 2.5. Scheduling and location problems

The ScheLoc problem combines the facility location problem and job scheduling problem and has gained more attention in operation research in recent years. ScheLoc problem was first studied by Hennes and Hamacher (2002) where a single machine was to be located in a network. Single machine ScheLoc problem was also investigated in Elvikis et al. (2009), Kalsch and Dreznér (2010), Akbarinasaji and Mckendall (2017). The problem was then extended to parallel machine ScheLoc problem and attracted many researchers. Heßler and Deghdak (2017) introduced the discrete parallel machine makespan (DPMM) ScheLoc problem in which multiple machines could be located to a finite set of candidate locations. The objective was to minimize the makespan. They proposed an integer programming formulation, four lower bounds, and several clustering heuristics improved by local search procedures. Wang et al. (2020b) studied the DPMM ScheLoc problem and built a MILP formulation based on the network-flow problem. They proposed two formulation-based heuristics to solve small-scale instances. For large-scale instances, a polynomial location–density–longest processing time (LDL) algorithm was developed. The proposed heuristics were verified to be competitive through computational experiments. Liu et al. (2019) considered the stochastic parallel machine ScheLoc problem with uncertain job processing times. A two-stage stochastic programming formulation was proposed to minimize the weighted sum of the location cost and the expectation of the total completion time. They developed an SAA method, a scenario-based heuristic, and a genetic algorithm for the problem.

## 2.6. Logic-based Benders decomposition and its applications

Benders decomposition method (Benders 1962) is effective in solving large-scale MIP problems. Its basic idea is to decompose the original problem into a relaxed master problem (MP) and a



linear subproblem (SP). The classic Benders decomposition (CBD) applies an iterative manner between MP and SP. Benders cuts are generated by solving the SP and are added to the MP in the subsequent iterations to obtain better solutions. The algorithm terminates when the global optimal solution is obtained or the time limit is reached. Hooker and Ottosson (2003) introduced the LBBD method, which generalizes the CBD by allowing the SP to take any form, instead of a linear program. LBBD has shown its superiority in various problems, including order acceptance and scheduling and operating room scheduling. Next, we briefly review the successful application of LBBD on these problems. Roshanaei et al. (2020) investigated the balanced DORS that considered two levels of balancing decisions. The problem was first formulated as a mixed-integer nonlinear program and then adapted into three variants by various reformulation-linearization techniques. A uni- and a bi-level LBBD were developed to solve the models. Guo et al. (2021) extended the DORS by considering stochastic surgery durations. The problem was modeled as a two-stage stochastic integer program and was reformulated via sample average approximation. The objective was to minimize the total operational and expected cancellation costs computed by subtracting the patient scheduling benefits from the sum of the surgical suite opening cost, the operating room opening cost, the postponement penalty cost, and the expected cancellation cost. Several decomposition schemes were developed for the problem. The first approach was a two-stage decomposition using classic Benders cuts and LBBD cuts. As an alternative to the two-stage one, a three-stage decomposition approach was proposed in which the subproblems were solved by a two-stage decomposition method.

Another successful application of LBBD comes from the order acceptance and scheduling (OAS) problem. Naderi and Roshanaei (2020) studied the OAS problem on identical parallel machines that simultaneously optimized the order acceptance, assignment, and scheduling. A MILP model was formulated and further enhanced by pre-processing techniques, valid inequalities, and dominance rules. The objective was to maximize the total profit, i.e., the difference between the revenue of accepted orders and the total tardiness cost. Based on the model structure, a CBD, an LBBD, and a Branch-Relax-and-Check (BRC) were proposed, respectively. The BRC, compared to the LBBD, was novel in that it incorporated temporary Benders cuts and designed problem-specific SP relaxations.

Some other recent applications include quay crane scheduling problem (Sun et al. 2019), home care scheduling (Grenouilleau et al. 2020), and supply chain network design (Naderi et al. 2020).

The novelty of this paper is multifold. First, we study a new MCVAS problem that jointly optimizes the number of vaccination sites to open, the locations of the opened vaccination sites, the accepted appointments, the assignments of appointments to open vaccination sites, and the schedule of vaccination service at each site. The formulated model is novel as none of the existing ones can

directly apply to our problem. Second, the proposed model and algorithm can efficiently provide solutions to practical COVID-19 vaccination programs. Third, we conduct numerical experiments and sensitivity analysis to draw managerial implications. Moreover, we can adjust the weights of different cost components in the objective function to better suit real-world situations or the decision-makers' preferences, making it applicable to different scenarios.

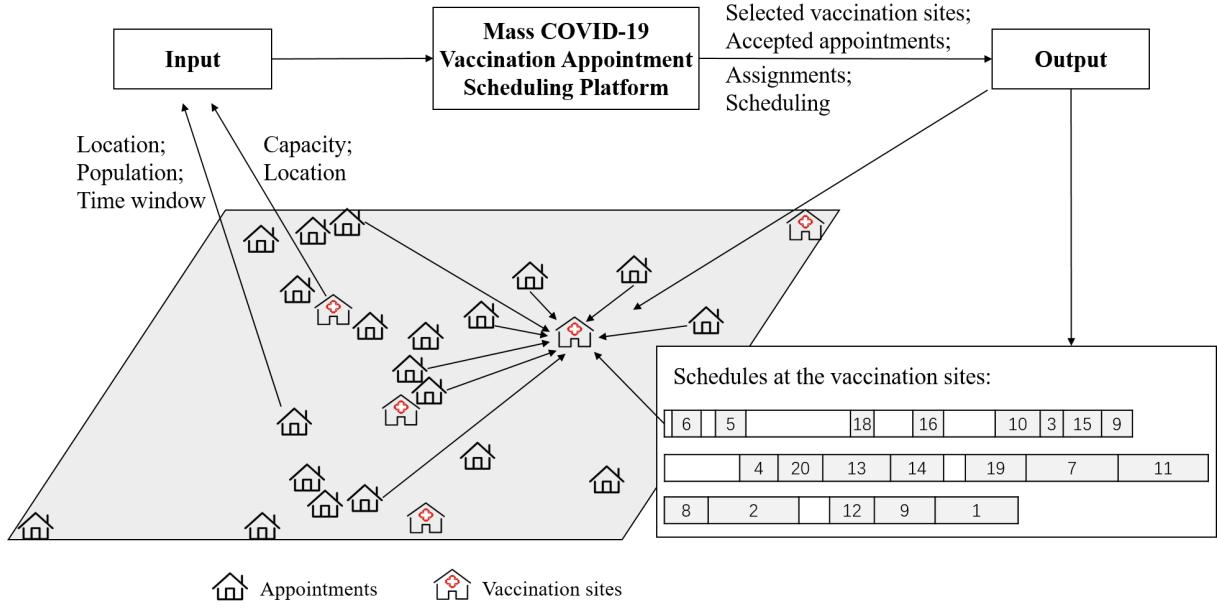
### 3. Problem description and formulation

In this section, we describe the problem formally at first and then propose the linear ordering formulation adapted from similar scheduling problems.

#### 3.1. Problem description

Suppose that during the mass COVID-19 vaccination, decision-makers have launched an appointment platform through which vaccine recipients can make appointments for vaccination. Due to the limited capacity of vaccine supply, personnel like healthcare workers are prioritized to be vaccinated. In this case, the platform receives appointments by groups. Each group is characterized by its number of recipients, geographical location, and self-imposed vaccination time window. The accepted appointments must be assigned to one of the available vaccination sites. Several critical decisions must be made, including which vaccination site to open, which appointment to accept, which site to assign the accepted appointments to, and each appointment's time slot.

Given a set  $K = \{1, 2, \dots, l\}$  of qualified vaccination sites, each of which has a fixed operating cost  $c_k$  ( $k \in K$ ) once opened, a specific geographical location with a coordinate  $(X_k, Y_k)$ , and a specific workload limit  $Q$ . A set  $J = \{1, 2, \dots, n\}$  of appointments are made. Each appointment  $j$  ( $j \in J$ ) is characterized with a non-negative processing time  $p_j$  which is positively correlated to the number of vaccine recipients in the group, and a time window  $(e_j, l_j)$  indicating the earliest start time and latest completion time. We consider half-soft time windows where  $l_j$  can be violated with a penalty cost  $q_j$  and the tardiness  $T_j$ .  $q_j$  is regarded as the weight of appointment  $j$  and is positively related to the population. The service start time  $e_j$  must be met, since otherwise, the vaccine recipients can hardly arrive at the site, which may cause a no-show, bringing unnecessary medical and human resource waste. Due to the workload limitation at each vaccination site, the appointments may sometimes be rejected or re-arranged to some other workdays. If an appointment  $j$  is rejected, a rejection cost  $r_j$  is incurred. The distance of the vaccine recipient group  $j$  and the vaccination site  $k$  is  $d_{jk}$ . Let  $S_j$  be the service start time of appointment  $j$ , its tardiness can be computed as  $T_j = \max(S_j + p_j - l_j, 0)$ . The objective of the mass COVID-19 vaccination appointment scheduling problem is to minimize the weighted sum of total fixed costs for opening vaccination sites, total travel distance of vaccination recipients, total rejection costs, and total weighted tardiness costs.



**Figure 1** Illustrative example

### 3.2. Illustrative example

To better understand the problem, we present an illustrative example of 20 appointments and five vaccination sites that are randomly dispersed on the planar in Figure 1. The mass COVID-19 vaccination appointment scheduling platform is the information collecting and processing center where the decision-making will determine the optimal solutions based on the input information. The input information covers the capacity and the geographical location of vaccination sites, the geographical location of recipients, the number of vaccine recipients, and the self-imposed time window. After the decision-making process, the output gives the solutions of the selected vaccination sites, the accepted appointments, the assignments, and the schedules at each vaccination site. The solution shows that all 20 appointments are accepted, and three of the five vaccination sites are selected. The schedules at the selected vaccination sites are also reported in the figure.

### 3.3. Linear ordering formulation

We formulate our problem as a MILP model based on linear ordering formulation, which performs well in scheduling problems (E. Dyer and Wolsey 1990). To this end, we define the following variables:

- $w_k$  equal to 1 if vaccination site  $k$  is open;
- $v_{jk}$  equal to 1 if appointment  $j$  is assigned to the vaccination site  $k$ ;
- $x_{ij}$  equal to 1 if appointment  $i$  precedes appointment  $j$ ;
- $z_j$  equal to 1 if appointment  $j$  is rejected;
- $S_j$  service start time of appointment  $j$ ;

$T_j$  Tardiness of appointment  $j$ .

The studied MCVAS problem can be formulated using the linear ordering (LO) model presented as follows:

$$\min \quad \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 + \theta_4 f_4 \quad (1)$$

s.t.

$$f_1 = \sum_{k \in K} c_k w_k \quad (2)$$

$$f_2 = \sum_{j \in J} \sum_{k \in K} q_j d_{jk} v_{jk} \quad (3)$$

$$f_3 = \sum_{j \in J} r_j z_j \quad (4)$$

$$f_4 = \sum_{j \in J} q_j T_j \quad (5)$$

$$\sum_{k \in K} v_{jk} + z_j = 1 \quad \forall j \in J \quad (6)$$

$$v_{jk} \leq w_k \quad \forall j \in J, k \in K \quad (7)$$

$$\sum_{j \in J} p_j v_{jk} \leq Q \quad \forall k \in K \quad (8)$$

$$S_j \geq e_j(1 - z_j) \quad \forall j \in J \quad (9)$$

$$S_j \geq S_i + p_i - L(3 - x_{ij} - v_{jk} - v_{ik}) \quad \forall i, j \in J, i < j, k \in K \quad (10)$$

$$S_i \geq S_j + p_j - L(2 + x_{ij} - v_{jk} - v_{ik}) \quad \forall i, j \in J, i < j, k \in K \quad (11)$$

$$T_j \geq S_j + p_j - l_j \quad \forall j \in J \quad (12)$$

$$T_j \geq 0 \quad \forall j \in J \quad (13)$$

$$w_k \in \{0, 1\} \quad \forall k \in K \quad (14)$$

$$z_j \in \{0, 1\} \quad \forall j \in J \quad (15)$$

$$v_{jk} \in \{0, 1\} \quad \forall j \in J, k \in K \quad (16)$$

$$x_{ij} \in \{0, 1\} \quad \forall i, j \in J, i < j, \quad (17)$$

where  $L$  is a large enough positive number and  $\theta_1, \theta_2, \theta_3$ , and  $\theta_4$  are the weights of the four components in the objective function (1) which aims at minimizing the weighted sum of four different objectives elaborated below. Objective (2) represents the total operational cost for opening vaccination sites. Objective (3) is the total weighted sum of the traveling distance of all the vaccination recipients in each group. Objective (4) is the total rejection costs of appointments been rejected. Objective (5) is the total weighted tardiness.

Constraints (6) ensure that each appointment is either accepted and assigned to a vaccination site or rejected. Constraints (7) indicate that an appointment can only be assigned to an open

vaccination site. Constraints (8) correspond to the working time limit constraints at each vaccination site, indicating that the total processing time of the appointments assigned to each site can not exceed the upper bound  $Q$ . Constraints (9) state that the service start time of appointment  $j$  is no early than its earliest start time  $e_j$ . Constraints (10) and (11) are the processing sequences constraints, indicating that if appointment  $j$  is served after appointment  $i$  at the same vaccination site, the service start time of appointment  $j$  must be greater than or equal to the completion time of appointment  $i$ . In real-world vaccination sites, social distance policy is strictly obeyed. The recipients cannot enter the vaccination zone until the former appointed group finishes the vaccination. It is also significant to avoid potential cross-infection. Constraints (12) and (13) compute the tardiness of each appointment. If the completion time completes before the due date, no tardiness is incurred. Otherwise, if the due date is violated, the tardiness will be counted by the difference between the service completion time and the due date. Constraints (14) – (17) are the variable domains.

### 3.4. Dominance rules and valid inequalities

We next propose some dominance rules (DR) and valid inequalities to improve the formulation.

**DR 1:** Appointment  $j$  must be rejected if  $\theta_2 q_j d_j^* + \theta_4 \max(e_j + p_j - l_j, 0) \geq \theta_3 r_j$ , where  $d_j^* = \min_{k \in K} d_{jk}$ .

**Proof:** We consider a solution in which appointment  $j$  satisfying the above condition is accepted. We only need to show that the objective value increases by removing order  $j$ . To this end, let us reject  $j$ , and keep other appointments unchanged. The objective function value will increase by the weighted rejection cost  $\theta_3 r_j$ , and at the same time decrease at least  $\theta_2 q_j d_j^* + \theta_4 \max(e_j + p_j - l_j, 0)$ , in which the first part is the weighted distance cost and the second is the lower bound of the total tardiness penalty cost. Therefore, if  $\theta_2 q_j d_j^* + \theta_4 \max(e_j + p_j - l_j, 0) \geq \theta_3 r_j$  stands for appointment  $j$ , the total cost will decrease when  $j$  is rejected. Therefore, it is beneficial to reject  $j$ .  $\square$

**DR 2:** Appointment  $j$  dominates  $i$  if  $p_j \leq p_i$ ,  $q_j \leq q_i$ ,  $e_j \leq e_i$ ,  $l_j \geq l_i$ ,  $r_j \geq r_i$ , and  $d_{jk} \leq d_{ik}, \forall k \in K$ .

**Proof:** This dominance rule indicates that if appointment  $j$  is rejected, appointment  $i$  will definitely be rejected, i.e.,  $z_j \leq z_i$  if they satisfy the above mentioned conditions. To prove the validity of **DR 2**, we prove that if appointment  $i$  is accepted and appointment  $i$  is rejected in a solution, we can obtain a better solution if we reject  $i$  and assign appointment  $j$  to the position of  $i$ . First, we consider the situation in which appointment  $i$  is accepted and assigned to vaccination site  $k$ , while appointment  $j$  is rejected. We assume that the finishing time of the appointments before appointment  $i$ , in an optimal solution, is  $T^-$ , which means that the start time of appointment  $i$  is  $S_i = \max(T^-, e_i)$ . Thus, the objective value related to appointment  $i$ , i.e.,  $F_i$ , can be presented as:

$$F_i = \theta_1 c_k w_k + \theta_2 q_i d_{ik} + \theta_3 r_j + \theta_4 \max(S_i + p_i - l_i, 0) \quad (18)$$

Then we consider another situation in which appointment  $j$  is accepted and assigned to vaccination site  $k$ , while appointment  $i$  is rejected. We schedule appointment  $j$  to exactly the position of appointment  $i$  in the first solution and keep all the other appointments unchanged. Similarly, we can compute  $F_j$  as:

$$F_j = \theta_1 c_k w_k + \theta_2 q_j d_{jk} + \theta_3 r_i + \theta_4 \max(S_j + p_j - l_j, 0) \quad (19)$$

The difference between  $F_i$  and  $F_j$  as:

$$F_j - F_i = \theta_2 (q_j d_{jk} - q_i d_{ik}) + \theta_3 (r_i - r_j) + \theta_4 (\max(S_j + p_j - l_j, 0) - \max(S_i + p_i - l_i, 0)) \quad (20)$$

As appointment  $i$  and  $j$  satisfy the conditions in **DR 2** and both the two appointments share the same  $T^-$ , we can obtain that  $S_j \leq S_i$ . We next prove  $F_j - F_i \leq 0$  with two cases:

**Case 1:**  $S_j + p_j - l_j \leq 0$ ,  $S_i + p_i - l_i > 0$ . Then equality (20) can be transformed into:

$$F_j - F_i = \theta_2 (q_j d_{jk} - q_i d_{ik}) + \theta_3 (r_i - r_j) - \theta_4 (S_i + p_i - l_i) < 0 \quad (21)$$

**Case 2:**  $S_j + p_j - l_j > 0$ ,  $S_i + p_i - l_i > 0$ . Then equality (20) can be transformed into:

$$F_j - F_i = \theta_2 (q_j d_{jk} - q_i d_{ik}) + \theta_3 (r_i - r_j) + \theta_4 (S_j - S_i + p_j - p_i + l_i - l_j) \leq 0 \quad (22)$$

Therefore, we can conclude that dominance rule **DR 2** is valid.  $\square$

Based on the above two dominance rules, we can add the following valid inequalities to the model.

$$z_j = 1 \quad \forall j \in J, \theta_2 q_j \min_{k \in K} d_{jk} + \theta_4 \max(e_j + p_j - l_j, 0) \geq \theta_3 r_j \quad (23)$$

$$z_j \leq z_i \quad \forall i, j \in J, p_j \leq p_i, q_j \leq q_i, e_j \leq e_i, l_j \geq l_i, r_j \geq r_i, d_{jk} \leq d_{ik} (k \in K). \quad (24)$$

## 4. Solution methods

The proposed MCVAS problem comprises a facility location problem and a multi-provider appointment scheduling problem with rejection. The facility location problem is proved to be NP-hard (Korupolu et al. 2000). The appointment scheduling problem can also be classified as NP-hard as a similar simplified single machine scheduling problem with rejection has already been proved to be NP-hard (Zhang et al. 2009). That means that our problem is also NP-hard. We propose an exact algorithm based on logic-based Benders decomposition (LBBD) to solve the problem.

#### 4.1. Logic-based Benders decomposition

The proposed LBB method decomposes the original MCVAS problem into an MP and a series of SPs. The MP tackles the vaccination site location, the appointment acceptance, and the assignment of accepted appointments to the vaccination sites. The SPs, based on the solutions of the MP, determine the appointment sequencing decisions at each vaccination site. The logic-based Benders cuts are generated by solving the SPs and are incorporated into the MP in the next iteration. This procedure is repeated until the optimal solution is obtained or the time limit is reached. Next, we present the MP, SP, and Benders cuts, respectively.

##### 4.1.1. Master problem

The MP is a relaxation of the original problem by fixing certain variables. Among the four objectives, we fix the total weighted tardiness by defining  $\Psi$  as the lower bound of the total weighted tardiness of all the accepted appointments. Thus, the MP can be formulated as follows:

$$\min \quad \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 + \theta_4 \Psi \quad (25)$$

s.t.

$$(2) - (4), (6) - (8), (14) - (16), (23), (24), \text{ and to} \\ CUTS. \quad (26)$$

The objective function of MP is the minimization of the former three objectives and the lower bound approximation of total weighted tardiness. Cuts (26) are added to MP iteratively to improve the value of  $\Psi$ . However, the lower bound is considerably weak, leading to lower computational efficiency. Next, we propose several valid inequalities to improve the lower bound of total weighted tardiness and computational efficiency.

##### 4.1.2. Strengthen the MP

We define  $\Phi_k$  as the total weighted tardiness of appointments at vaccination site  $k$ , i.e.,  $\Phi_k = \sum_{j \in J} v_{jk} q_j T_j, \forall k \in K$ . Then, we can add the following inequalities to help the MP obtain a better lower bound.

$$\Psi \geq \sum_{k \in K} \Phi_k \quad (27)$$

$$\Psi \geq \sum_{j \in J} q_j T_j \quad (28)$$

$$\Phi_k \geq (\min_{j \in J} q_j) \left( \sum_{j \in J} p_j v_{jk} + \min_{j \in J} (e_j) - \max_{j \in J} (l_j) \right) \quad \forall k \in K \quad (29)$$

$$T_j \geq e_j + p_j - l_j \quad \forall j \in J \quad (30)$$

$$\Phi_k \geq 0 \quad \forall k \in K \quad (31)$$

$$T_j \geq 0 \quad \forall j \in J. \quad (32)$$

Constraint (27) indicates that the total weighted tardiness is no less than the sum of weighted tardiness generated at each vaccination site. Constraint (28) means that the total weighted tardiness is at least the sum of weighted tardiness of all appointments. Inequalities (29) indicate that the total weighted tardiness generated at vaccination site  $k$  is at least the minimum potential weighted tardiness of the last-served appointment. We assume that the appointment has the minimum weight  $\min_{j \in J} q_j$ , the minimum start time  $\min_{j \in J} e_j$ , and the largest due date  $\max_{j \in J} l_j$ . Inequalities (30) ensure that if an appointment  $j$  is assigned to vaccination site  $k$ , its tardiness is at least the total of its earliest start time and processing time, and minus its completion due date. The SP can be formulated and solved as follows upon solving the MP.

#### 4.1.3. Subproblems

After the MP is solved, the value of variables  $w_k$ ,  $z_j$ , and  $v_{jk}$  ( $j \in J, k \in K$ ), indicating the vaccination site selection, appointment acceptance, and appointment- vaccination site assignments, are all fixed. Therefore, the SP is several single machine scheduling problems dealing with the vaccination sequences of assigned appointments with time windows to minimize the total weighted tardiness at each open vaccination site. We define a subset  $J_k$  of the appointments that are assigned to vaccination site  $k$ , i.e.,  $J_k = \{j | v_{jk} = 1, \forall j \in J\}$ . We define a binary sequencing variable  $y_{ij}$  ( $i, j \in J_k$ ) taking value 1 if vaccine recipients of appointment  $j$  are vaccinated immediately after the recipients of appointment  $i$ . Then, the SP for vaccination site  $k$  ( $SP_k$ ) can be formulated as follows:

$$SP_k: \quad \min \sum_{j \in J_k} q_j T_j \quad (33)$$

s.t.

$$y_{ij} + y_{ji} \leq 1 \quad \forall i, j \in J_k \quad (34)$$

$$S_j \geq S_i + p_i - L(1 - y_{ij}) \quad \forall i, j \in J_k \quad (35)$$

$$S_j \geq e_j(1 - z_j) \quad \forall j \in J_k \quad (36)$$

$$T_j \geq S_j + p_j - l_j \quad \forall j \in J_k \quad (37)$$

$$T_j \geq 0 \quad \forall j \in J_k \quad (38)$$

$$y_{ij} \in \{0, 1\} \quad \forall i, j \in J_k. \quad (39)$$

The objective (33) minimizes the total weighted tardiness of all the accepted appointments that are assigned to the vaccination site  $k$ . Constraints (34) avoid an appointment  $j$  being the predecessor and successor of another one simultaneously. Constraints (35) and (36) are the vaccination completion time constraints for two adjacent appointments, indicating that the start time of an appointment can neither be early than the completion time of its predecessor nor be early than its own earliest start time. Constraints (37) compute tardiness of each assigned appointments.



The single machine scheduling problem with release time and due date to minimize total weighted tardiness is denoted as  $1|e_j|\sum q_j T_j$  according to the standard three-field notation method by Graham et al. (1979). The  $1|e_j|\sum q_j T_j$  problem is NP-hard (Pinedo and Rammouz 1988) and has been investigated by many researchers using different approaches (França et al. 2001, Cheng et al. 2005, Cordone and Hosteins 2019). Among the exact algorithms for the  $1|e_j|\sum q_j T_j$  problem, the dynamic programming (DP) method proposed by Tanaka and Fujikuma (2012) is proved to be efficient. The DP algorithm has been successfully applied in other problems in Tanaka and Sato (2013), Tanaka and Araki (2013), and Şen and Bülbül (2015). We apply the DP to optimally solve the  $SP_k$ , which is exactly a  $1|e_j|\sum q_j T_j$  problem. Once a solution to the SP is obtained, Benders cuts are generated.

#### 4.1.4. Benders cuts

After the  $SP_k$  is solved, the value of the total tardiness generated at each opened vaccination site  $k \in K$  is fixed, denoted by  $\Phi_k^*$ . We propose a generic Benders optimality cut, which is shown as follows:

$$\Phi_k \geq \Phi_k^* \left( \sum_{j \in J_k} v_{jk} - |J_k| + 1 \right) \quad \forall k \in K. \quad (40)$$

Cuts (40) indicate that in subsequent iterations, if the same set of appointments  $J_k$ , probably among other appointments, are assigned to vaccination site  $k$ , the total tardiness at vaccination site  $k$  is at least  $\Phi_k^*$ . Note that if any appointment  $j'$  in set  $J_k$  changes, i.e.,  $\exists j' \in J_k, v_{j'k} = 0$ , in the following iterations, the cuts (40) become non-binding. The cuts (40) are dynamically generated and added to the MP to drive the MP and SP converges to optimality. The process terminates if the optimal solution is captured or the time limit is reached.

**Theorem 1.** Optimality cut (40) is valid.

**Proof.** Let the set  $J_k$  contains all the appointments assigned to vaccination site  $k$ , i.e.,  $J_k = \{j | v_{jk} = 1, \forall j \in J\}$ . Let set  $J_k^h$  be the set containing all the appointments that are assigned to vaccination site  $k$  in the subsequent iteration  $h$ . Thus, we consider two cases.

**Case 1:**  $J_k^h \cap J_k = J_k$ . This case means that in subsequent iteration  $h$ , all the appointments in set  $J_k$  are assigned to vaccination site  $k$ , i.e.,  $\sum_{j \in J_k} v_{jk} = |J_k|$ . By adding this into cut (40) we can obtain  $\Phi_k \geq \Phi_k^*$ , which indicates that the cut is valid.

**Case 2:**  $J_k^h \cap J_k \neq J_k$ . This case means that at least one appointment  $j^* \in J_k$  is removed in the set  $J_k^h$ , i.e.,  $v_{j^*k} = 0$ . Therefore we can obtain  $\sum_{j \in J_k} v_{jk} < |J_k|$ . Thus the right side of cut (40) becomes non-positive and we can obtain  $\Phi_k \geq 0$  based on constraint (31). This indicate that the cut (40) will not remove new feasible solutions in the subsequent iterations.

---

**Algorithm 1** LBB method for the MCVAS problem

---

1. Formulate the MP presented in Section 4.1.1
  2. Strengthen the MP using inequalities introduced in Section 4.1.2
  3. Solve MP with the B&C procedure in a single search tree, during the branching process:
  4. **If** an integer solution to MP is obtained, **do**
  5.   Get the value  $v_{jk}$  and the set  $J_k$  of appointments assigned to  $k$
  6.   Solve the  $SP_k$  for each  $J_k$  using the DP introduced in Section 4.1.3
  7.   Generate Benders cuts (40)
  8.   Add the cuts (40) to the MP, and **go to** Step 3
  9. **End if**
  10. Output the best obtained solution
- 

To conclude, the cut will limit the total tardiness  $\Phi_k$  to at least  $\Phi_k^*$  when the same set of appointments as  $J_k$  are assigned to vaccination site  $k$  in the subsequent iterations and will not remove new feasible solutions. Thus the cut (40) is valid. The cut (40) is similar with some existing papers e.g., scheduling problem by Zhang et al. (2021) and OAS problem by Naderi and Roshanaei (2020).

#### 4.1.5. Outline of the LBB method

The LBB method first formulates the MP by relaxing the binary sequencing variables  $x_{ij}$ . Then the MP is solved by a branch-&-cut (B&C) method using a single search tree with the implementation of *branch and check*. At each node where a feasible solution to the MP is identified, the corresponding SP is solved with given appointment assignments. The solution to the SP is then used to generate Benders cuts, which are added to the MP. The MP branching process continues with added Benders cuts. This procedure is repeated until an optimal solution is obtained or a time limit is reached. The outline of the LBB method is shown in Algorithm 1.

#### 4.2. A matheuristic with predetermined sequences

This subsection proposes a matheuristic (MH) based on predetermined appointment sequences for the studied MCVAS problem. The basic idea of a matheuristic is to combine mathematical programming models with heuristic methods (Maniezzo et al. 2010). This kind of hybridization enables a matheuristic to obtain near-optimal solutions efficiently by capturing the property and characteristics of a problem, which is essential in solving large-scale optimization problems. Matheuristics have been successfully applied for solving complex combinatorial optimization problems, including flow-shop scheduling (Ta et al. 2015), knapsack problem (Lahyani et al. 2019), vehicle routing problem (Wang et al. 2017), and parallel machine scheduling problem (Fanjul-Peyro et al. 2017, Dang et al. 2020).

For the studied MCVAS, we observe that the main difficulty of the model comes from the scheduling part, i.e., sequencing appointments at each vaccination site. This corresponds to solving a series

of strongly NP-hard  $1|e_j|\sum q_j T_j$  problems. To reduce the complexity of the MCVAS problem yet retain most information (constraints), we develop a matheuristic MH for the MCVAS problem. The MH first sequence all appointments at each vaccination site using some priority rules. Then an approximate MILP model is solved with predetermined appointment service sequences. We have performed preliminary experiments to select the best sequencing rules. The tested rules include sequencing the appointments using the earliest start time first rule, earliest due date first rule, longest processing time first rule, minimum  $(l_j - e_j - p_j)$  first rule, and minimum  $(l_j + e_j)$  rule. Preliminary results show that sequencing appointments using the small value of  $(l_j + e_j)$  first rule leads to a better solution. Thus we use this rule to sort all appointments.

We define a set  $A$  of appointments, which contains all appointments in  $J$  while the appointments in it are sorted in non-decreasing order of  $(l_j + e_j)$ . Let  $S'_{jk}$  be the service start time of appointment  $j \in A$  at vaccination site  $k$ . The approximate MILP model (AP) can be formulated as follows:

$$\min \quad \theta_1 f_1 + \theta_2 f_2 + \theta_3 f_3 + \theta_4 f_4 \quad (41)$$

s.t.

(2) – (8), (12) – (16), (23), (24), and to

$$S_j \geq S'_{jk} - L(1 - v_{jk}) \quad \forall j \in A, k \in K \quad (42)$$

$$S'_{jk} \geq e_j v_{jk} \quad \forall j \in A, k \in K \quad (43)$$

$$S'_{jk} \geq S'_{j-1,k} + p_{j-1} v_{j-1,k} \quad \forall j \in A, k \in K. \quad (44)$$

Constraints (42) indicate that the start time of appointment  $j$  is at least the start time at vaccination site  $k$  if it is assigned to it. Constraints (43) make sure that the start time of appointment  $j$  at vaccination site  $k$  is no early than the earliest start time  $e_j$  if the appointment is accepted and assigned to vaccination site  $k$ . Constraints (44) provide the sequencing rule of two successive appointments by the predefined sequences. Specifically, the start time of appointment  $j$  at vaccination site  $k$  is greater than or equal to the sum of the start time and processing time of its predetermined predecessor  $j - 1$ , if  $j - 1$  is assigned to the same vaccination site. If appointment  $j - 1$  is not assigned, the start time of  $j$  on vaccination site  $k$  is the start time of  $j - 1$ . That means that the recipients of an appointment have to wait to complete the former appointment before entering the vaccination zone and starting to be vaccinated. By defining constraints (44), we can omit the sequencing variable  $x_{ij}$  and the corresponding constraints in the LO model.

MH solves the above model to obtain near-optimal solutions to the original MCVAS problem. These solutions are then improved by solving a series of  $1|e_j|\sum q_j T_j$  problems, one for each vaccination site, using the DP of Tanaka and Fujikuma (2012). The detailed steps of the MH is presented in Algorithm 2.

---

**Algorithm 2** MH method for the MCVAS problem

---

1. Define a set  $A$  of appointments, where  $A = J$
  2. Sort the appointments in  $A$  in non-decreasing order of  $(l_j + e_j)$
  3. Formulate the approximate model AP with the set  $A$
  4. Solve the model AP using an off-the-shelf solver
  5. Get the value of  $w_k$ ,  $v_{jk}$ ,  $z_j$ , and  $J_k = \{j | w_k = 1, v_{jk} = 1\}$
  6. **For** each  $J_k \neq \emptyset$ , **do**
  7.   Solve  $1|e_j| \sum q_j T_j$  problem
  8.   Get the value of  $T_j$
  9. **End for**
  10. Output the best obtained solution
- 

## 5. Computational study

In this section, we conduct numerical experiments to evaluate the performances of the proposed model, the exact LBBD method, and the matheuristic MH. We next detail the data generation schemes in Section 5.1. We then report the computational results on small- to large-sized instances with up to 100 appointments and 20 vaccination sites in Sections 5.2–5.4. We further test practical-sized instances with up to 500 appointments and 50 vaccination sites in Section 5.5. Finally, we perform sensitivity analysis on different cost components and provide some insights.

All the models and algorithms are coded in C++ linked with CPLEX 12.10. For each run, the optimization process terminates when the optimal solution is obtained, i.e., the upper bound equals the lower bound, or the time limit is reached, which is set to be 3600 seconds in this paper. All experiments were run on a computer with an Intel Xeon CPU E5-2690 v3 at 2.60 GHz with 32 GB RAM.

### 5.1. Data generation

We generate 56 instances with up to 500 appointments and 50 vaccination sites. Each group may involve tens of even hundreds of individual recipients. Therefore, in our problem, 500 appointments may include tens of thousands of recipients. The 56 instances are divided into two groups. The first group contains 40 instances with up to 100 appointments and 20 vaccination sites, i.e.,  $n = \{10, 20, \dots, 100\}$  and  $l = \{5, 10, 15, 20\}$ . The second group contains practical-sized instances with up to 500 appointments and 50 vaccination sites, i.e.,  $n = \{200, 300, 400, 500\}$  and  $l = \{20, 30, 40, 50\}$ . For each combination of the number of  $n$  and  $l$ , we generate one instance, totaling 56 instances. The parameters are generated as follows. The coordinates of the vaccination site and the vaccine recipients are randomly generated in a  $200 \times 200$  plane, with the unit of 100 meters. The travel distance  $d_{jk}$  from recipient  $j$  to vaccination site  $k$  is calculated using the euclidean distance  $d_{jk} = \left\lceil \sqrt{(X_j - X_k)^2 + (Y_j - Y_k)^2} \right\rceil$ . The service duration  $p_j$  of appointment  $j$  is randomly generated from

the set  $p_j = \{30, 40, \dots, 120\}$ . The fixed cost  $c_k$  for opening a vaccination site is randomly generated from  $c_k = \{400, 410, \dots, 600\}$ . The maximum working time of vaccination sites is set to  $Q = 600$ , with a unit of minutes. We define two parameters  $center_j$  and  $width_j$  for each appointment  $j$ , the time window  $(e_j, l_j)$  is generated as follows:

$$\begin{aligned} center_j & \text{ randomly generated } [0, Q - p_j] \\ width_j & \text{ randomly generated } Q * [1/8, 1/4] \\ e_j & \max (center_j - width_j/2, 0) \\ l_j & \min (center_j + width_j/2 + p_j, Q). \end{aligned}$$

Each appointment  $j$  is associated with a specific weight  $q_j$  which is dependent on its service duration. The weight  $q_j$  for appointment  $j$  is set using formula (45). The rejection cost  $r_j$  is set to  $r_j = 100 * q_j + 100, \forall j \in J$ . Without loss of generality, we set the weights of the objectives  $\theta_1 = \theta_2 = \theta_3 = \theta_4 = 1$ .

$$q_j = \begin{cases} 1, & 30 \leq p_j \leq 50, \\ 2, & 51 \leq p_j \leq 80, \\ 3, & 81 \leq p_j \leq 120. \end{cases} \quad \forall j \in J \quad (45)$$

## 5.2. Computational results for small-sized instances

To better analyze the results, we divide the first group of instances into three subsets according to instance sizes: (i) small-sized instances with  $n = \{10, 20, 30, 40\}$ ; (ii) medium-sized instances with  $n = \{50, 60, 70\}$ ; and (iii) large-sized instances with  $n = \{80, 90, 100\}$ . We first show the results of the small-sized instances in Table 1. The first three columns show the serial number of the instance, the number of appointments, and the number of vaccination sites, respectively. The bracket (*opt*) in the last row indicates the number of optimally solved instances. The relative optimality gaps and computation times obtained by the model and solution methods are then reported for each method. The *Gap*(%) is calculated as follows:

$$Gap (\%) = \frac{UB - LB}{UB} \times 100 \quad (46)$$

where UB and LB are the upper and lower bounds, respectively. Since MH cannot provide lower bounds, it uses the best lower bounds of LO and LBBD to calculate its average gap with the upper bounds it obtained.

From Table 1, we can find that all the models and solution methods perform well for small-sized instances. Among them, LBBD optimally solves all 16 small-sized instances, better than LO and

**Table 1 Computational results for small-sized instances**

Instances			Gap (%)			Time (s)		
No.	$n$	$l$	LO	MH	LBBB	LO	MH	LBBB
1	10	5	0.00	0.00	0.00	0.16	0.06	0.06
2		10	0.00	2.01	0.00	0.17	0.05	0.25
3		15	0.00	0.00	0.00	0.16	0.08	0.09
4		20	0.00	0.00	0.00	0.34	0.14	0.14
5	20	5	0.00	2.41	0.00	1.62	0.23	0.25
6		10	0.00	1.33	0.00	1.36	0.12	0.22
7		15	0.00	0.00	0.00	0.30	0.11	0.09
8		20	0.00	0.00	0.00	2.85	0.22	0.53
9	30	5	0.00	0.00	0.00	35.18	0.72	0.56
10		10	0.00	0.00	0.00	2.93	0.31	0.23
11		15	0.00	0.00	0.00	710.80	12.18	17.27
12		20	1.84	0.00	0.00	3600.08	10.26	40.95
13	40	5	0.00	0.01	0.00	211.71	1.75	3.17
14		10	0.55	0.37	0.00	3600.07	50.40	20.40
15		15	0.00	0.37	0.00	354.81	13.10	4.43
16		20	0.54	0.38	0.00	3600.14	26.10	9.94
Average (opt)			0.18 (13)	0.43 (9)	0.00 (16)	757.67	7.24	6.16

MH. LO obtains 13 optimal solutions and obtains an average gap of 0.18%. MH is comparatively inferior as it solves only nine instances to optimality, and the average gap of MH is 0.43%. In terms of computational efficiency, MH is competitive with LBBB. They solve each small-sized instance in at most 50.4 seconds, and the average computation time is less than eight seconds. LO is much slower than LBBB and MH, with an average computation time of 757.67 seconds. It can be observed in Table 1 that the average computation time increases when the instances get larger, which generally applies for all three methods. For example, the average computation time of LO is 0.21s for instances with ten appointments. At the same time, it increases to 1087.25 seconds when the number of appointments is 30 and further increases to 1941.68 seconds for instances with 40 appointments. When the number of appointments is ten, the average computation time of both MH and LBBB is shorter than 0.2 seconds, but they increase to an average of 22.84 and 9.48 seconds for the 40-appointment instances, respectively. In summary, LBBB outperforms LO and MH as it obtains more optimal solutions in a shorter time. MH is inferior to LO in solution quality but performs better than it in computational efficiency.

### 5.3. Computational results for medium-sized instances

Table 2 shows the computational results for medium-sized instances. LBBB solves eight of the 12 medium-sized instances optimally. While LO only obtains one optimal solution, none is solved optimally by MH. The average gap of MH is 0.84%, being second to that of LBBB (0.42%), and is better than that of LO (2.03%). The average computation time of MH is 523.72 seconds, which

**Table 2 Computational results for medium-sized instances**

Instances			Gap (%)			Time (s)		
No.	$n$	$l$	LO	MH	LBBD	LO	MH	LBBD
17	50	5	1.66	0.68	0.00	3600.05	40.23	2363.46
18		10	2.49	0.13	0.00	3600.10	116.75	787.68
19		15	3.19	0.67	0.62	3600.13	876.27	3600.10
20		20	1.64	0.04	0.00	3600.16	63.37	666.44
21	60	5	1.32	1.54	1.28	3600.07	48.61	3600.05
22		10	0.46	0.71	0.00	3600.10	28.24	2095.51
23		15	5.60	1.69	1.82	3600.14	3600.05	3600.07
24		20	2.89	0.14	0.00	3600.18	950.11	560.84
25	70	5	1.56	0.93	1.27	3600.10	53.13	3600.05
26		10	0.87	0.49	0.00	3600.14	276.57	311.57
27		15	2.71	2.09	0.00	3600.16	187.51	236.95
28		20	0.00	0.95	0.00	1636.40	43.77	13.85
Average (opt)			2.01 (1)	0.84 (0)	0.42 (8)	3436.48	523.72	1786.38

**Table 3 Computational results for large-sized instances**

Instances			Gap (%)			Time (s)		
No.	$n$	$l$	LO	MH	LBBD	LO	MH	LBBD
29	80	5	0.18	0.15	0.00	3600.11	9.69	141.37
30		10	5.33	4.52	4.65	3600.13	3600.04	3600.04
31		15	3.67	1.67	2.52	3600.25	3600.21	3600.10
32		20	3.00	0.42	0.66	3600.38	3600.13	3600.10
33	90	5	0.29	0.02	0.00	3600.11	11.34	127.81
34		10	3.51	2.29	1.98	3600.25	3600.04	3600.05
35		15	8.62	6.73	8.17	3600.30	3600.05	3600.08
36		20	5.10	4.58	3.79	3600.35	3600.05	3600.19
37	100	5	0.14	0.00	0.00	3600.11	23.88	45.01
38		10	2.30	2.30	2.48	3600.35	3600.10	3600.11
39		15	0.88	0.01	0.00	3600.39	1961.31	1628.84
40		20	6.07	2.40	4.23	3600.58	3600.10	3600.08
Average (opt)			3.26 (0)	2.09 (1)	2.37 (4)	3600.28	2567.25	2561.98

is a significant advantage over both LBBD and LO. That indicates that MH is good at finding high-efficiency near-optimal solutions, even though its performance in obtaining optimal solutions is not as good as LBBD. The average computation time of LBBD and LO are 1786.38 and 3436.48 seconds, respectively.

#### 5.4. Computational results for large-sized instances

The computational results for large-sized instances are reported in Table 3. The table shows that the large-sized instances are more difficult to solve for all the model and solution methods. MH and LBBD are competitive in solution quality and computational efficiency. MH obtains an average gap of 2.09% and an average computation time of 2567.25s, while LBBD is 2.37% and 2561.98s, respectively. However, LBBD solves more instances to optimality than MH (4 VS 1). LO solves no

**Table 4** Computational results for practical-sized instances

Instances			Gap (%)			Time (s)		
No.	$n$	$l$	LO	MH	LBB	LO	MH	LBB
1	200	20	13.40	4.24	4.77	3601.77	3600.30	3600.21
2		30	30.49	6.41	13.13	3602.78	3600.53	3600.36
3		40	14.71	7.65	11.94	3603.73	3600.39	3600.47
4		50	32.11	8.68	21.90	3605.62	3600.32	3600.44
5	300	20	48.01	3.26	4.31	3603.40	3600.36	3600.30
6		30	16.03	4.07	3.86	3606.17	3600.53	3600.43
7		40	30.62	9.03	19.07	3606.45	3600.86	3608.41
8		50	87.50	9.12	19.48	3610.97	3600.75	3600.32
9	400	20	-	0.80	1.04	-	3600.50	3600.22
10		30	45.37	2.83	4.96	3611.31	3600.92	3601.36
11		40	-	5.95	10.15	-	3601.06	3600.60
12		50	89.29	16.62	18.77	3615.42	3601.11	3601.95
13	500	20	-	1.01	1.59	-	3600.88	3607.73
14		30	-	2.28	4.15	-	3600.99	3600.38
15		40	-	5.10	7.61	-	3601.36	3617.77
16		50	-	5.21	10.50	-	3601.50	3605.73
Average			-	5.77	9.83	-	3600.77	3602.92

instance optimally, and both the average gap (3.26%) and the average computation time (3600.28) are inferior to the other two approaches.

### 5.5. Computational results for practical-sized instances

In this subsection, We further explore the performances of LO, MH, and LBB on the practical-sized instances with up to 500 appointments and 50 vaccination sites. The results are shown in Table 4. The symbol “-” indicates that the approach obtains no feasible solution. From the table, we can find that LO fails to solve six of the eight instances with 400 and 500 appointments, while MH and LBB can provide feasible solutions for all the practical-sized instances. None of the instances is optimally solved by all three approaches. However, MH outperforms LBB and LO, with the lowest average gap of 5.77%. The average gap of LBB is 9.83%. Both LBB and MH have obvious advantages compared with LO. Generally, the performance of LO decreases when the instances become large. For example, the average gap of LO is 22.68% for instances with 200 appointments, but it turns to be 45.54% when the instances have 300 appointments. LO does not solve half of the 400-appointment instances and all the 500-appointment instances. For both MH and LBB, the increase of the number of vaccination sites makes gaps grow simultaneously. For instances with 400 appointments, the gap of MH grows from 0.80% to 16.62% when the number of vaccinations increases from 20 to 50. At the same time, the gap of LBB grows from 1.04% to 18.77%. To conclude, the MH has a better performance than LO and LBB for practical-sized instances with hundreds of appointments and dozens of vaccination sites.



In summary, we obtain the following observations through analyzing the numerical results.

1. The MCVAS problem investigated in this paper is complicated as the LO formulation fails to provide satisfactory solutions efficiently. Though it obtains 13 out of 16 optimal solutions for small-sized instances, the average computational time is about 100 times longer than that by MH and LBBB.

2. The LBBB algorithm performs best on the instances in the first group with up to 100 appointments. In the 40 instances, LBBB obtains 28 optimal solutions, far more than that obtained by MH and LO (10 and 14, respectively). LBBB is also time-saving compared with LO.

3. The MH applies a simple but effective heuristic idea which shows some advantages in solving large-sized instances. The MH first solves a series of  $1|e_j|\sum q_j T_j$  problems to obtain the predetermined sequences of appointments at each site. Then a simplified approximated model is solved to provide near-optimal solutions. The computational results on large- and practical-sized instances show that it outperforms LO and LBBB in both average optimality gap and computation time.

## 5.6. Sensitivity analysis

In this subsection, we make some sensitivity analyses to compare the effects of each cost component in our problem. For each experiment, we choose one cost component and change its weight to  $\{0, 2, 3, 5\}$ , respectively, while fixing the other three weights by one.  $\theta_i = 0$  ( $i = 1, 2, 3, 4$ ) means that the specific cost component is not considered in the objective function. We select seven instances in the small- and medium-sized subsets with appointments  $n = \{10, 20, \dots, 70\}$  and with five vaccination sites. The first five columns are the weight value, the objective function value (Obj), the LB, the average gap, and computation time, respectively. To better observe the effects, we output the value of the total fixed location cost, the total weighted distance cost, the total rejection cost, and the total weighted tardiness of each solution, and denote them as FLC, DC, RC, and TC in the tables, respectively. We also count the number of accepted appointments, the number of selected vaccination sites, and the number of optimal solutions and denote them as AA, SVS, and OPT. The solutions are obtained by LBBB and compared with the original solution where all the weights are equal to one. Note that the values in each table are the average of the seven instances solutions.

### 5.6.1. Fixed location cost

From Table 5, we can conclude that with the increase of  $\theta_1$ , the cost paid on opening new vaccination sites gets much higher. The average number of vaccination sites decreases from 4.86 to 0.29 when the value of  $\theta_1$  increases from 0 to 5. At the same time, the number of accepted appointments decreases from 31.57 to 3.71. We see that most appointments are rejected when the value of  $\theta_1$  is large. In extreme situations with high fixed location costs, rejecting appointments seems more

**Table 5 Effects of fixed location cost**

$\theta_1$	Obj	LB	Gap (%)	Time (s)	FLC	DC	RC	TC	AA	SVS	OPT
0	5866.86	5823.78	0.43	2057.22	2388.57	2656.86	3142.86	67.14	31.57	4.86	3.00
1	7972.00	7924.67	0.36	1366.80	1947.14	2649.14	3342.86	32.86	31.00	4.00	5.00
2	9829.86	9768.53	0.40	1543.97	1641.43	2324.14	4128.57	94.29	28.86	3.43	4.00
3	11311.43	11271.56	0.23	1326.74	1315.71	1820.00	5485.71	58.57	24.71	2.86	6.00
5	12629.86	12629.86	0.00	2.92	117.14	215.57	11828.57	0.00	3.71	0.29	7.00

**Table 6 Effects of distance cost**

$\theta_2$	Obj	LB	Gap (%)	Time (s)	FLC	DC	RC	TC	AA	SVS	OPT
0	4708.57	4685.72	0.30	1087.02	1985.71	7065.14	2714.29	8.57	33.00	4.14	5.00
1	7972.00	7924.67	0.36	1366.80	1947.14	2649.14	3342.86	32.86	31.00	4.00	5.00
2	10130.57	10130.57	0.00	73.92	1742.86	1756.71	4842.86	31.43	26.14	3.57	7.00
3	11429.29	11429.29	0.00	0.71	1218.57	846.43	7642.86	28.57	16.57	2.57	7.00
5	12432.86	12432.86	0.00	0.12	608.57	175.43	10928.57	18.57	5.71	1.14	7.00

cost-saving than accepting them. Accepting means more fixed location costs, distance costs, and potentially tardiness costs. From this perspective, the problem becomes easier to solve as we find that the OPT increases from three to seven, and the average gap decreases from 0.43% to 0. The result shows that the fixed location cost is an essential factor in the vaccination appointment scheduling problem.

### 5.6.2. Distance cost

Table 6 shows the computational results when weight  $\theta_2$  increases from zero to five. A similar trend as it is in Table 5 can be observed. However, we can find that when distance cost is not considered, more vaccination sites are opened, and more recipients are accepted, resulting in lower rejection costs and tardiness costs. Meanwhile, the distance cost is so high that it is nearly three times when  $\theta_2 = 1$ . This means that some recipients might be assigned to vaccination sites that are very far away from them, which may cause dissatisfaction. When  $\theta_2 = 5$ , only a tiny portion of the appointments are accepted. High distance costs indicate the difficulty of traveling during the pandemic. In that case, the mass COVID-19 vaccination program will have a high risk of failure. This suggests that the program should be initiated with good transportation accessibility, and vaccination sites should be established at places that are not very geographically far from the appointments.

### 5.6.3. Rejection cost

The weight  $\theta_3$  is special compared with the former two. It is simply to understand that all appointments will be rejected if rejection cost is not considered in the objective function. Thus, when  $\theta_3 = 0$ , we add  $z_j = 0, \forall j \in J$  to the model and remove the **DR 1** and the working time limit of vaccination sites to make sure that no appointment is rejected. The computational results reported

**Table 7 Effects of rejection cost**

$\theta_3$	Obj	LB	Gap (%)	Time (s)	FLC	DC	RC	TC	AA	SVS	OPT
0	12608.00	6796.42	25.85	2058.23	2030.00	4459.43	0.00	6118.57	40.00	4.14	3.00
1	7972.00	7924.67	0.36	1366.80	1947.14	2649.14	3342.86	32.86	31.00	4.00	5.00
2	10909.43	10784.38	0.70	2057.28	2030.00	3162.29	2814.29	88.57	32.86	4.14	3.00
3	13735.29	13528.40	0.96	2057.30	2030.00	3263.86	2742.86	212.86	33.00	4.14	3.00
5	19153.29	18967.50	0.59	2057.30	2030.00	3367.57	2728.57	112.86	33.00	4.14	3.00

**Table 8 Effects of tardiness cost**

$\theta_4$	Obj	LB	Gap (%)	Time (s)	FLC	DC	RC	TC	AA	SVS	OPT
0	7687.86	7687.86	0.00	0.29	1947.14	2483.57	3257.14	934.29	31.14	4.00	7.00
1	7972.00	7924.67	0.36	1366.80	1947.14	2649.14	3342.86	32.86	31.00	4.00	5.00
2	8018.14	7924.33	0.72	1211.24	1947.14	2653.86	3371.43	22.86	30.86	4.00	5.00
3	8025.71	7932.80	0.72	1313.60	1947.14	2648.57	3357.14	24.29	30.86	4.00	5.00
5	8061.14	7934.20	0.97	1370.28	1947.14	2749.71	3314.29	10.00	30.86	4.00	5.00

in Table 7 show that the other three cost components are substantial. For example, the total tardiness penalty cost is 6118.57 when  $\theta_3 = 0$  but it decreases to 32.86 when  $\theta_3 = 1$  and appointment rejection is allowed. When the weight  $\theta_3$  of rejection cost gets higher, it becomes harder to reject any appointments. This might reflect the reality that the governments are desperate to complete the vaccination for all the citizens to accelerate the recovery of the regional economy. However, the increase in the rejection cost makes the problem much harder to solve. Table 7 shows that when  $\theta_3$  increases from one to two, the computation time sharply increases from 1366.80s to 2057.28s, and the average gap increases from 0.36 to around 0.70%. For the four cost components, the total rejection costs decrease while the other components increase because more appointments are accepted, and more vaccination sites are open.

#### 5.6.4. Tardiness cost

The last cost component is the total weighted tardiness, indicating the timeliness of the vaccination appointment scheduling. From Table 8, we can find that when  $\theta_4 = 0$ , the problem becomes much easier to solve as all the seven instances are solved to optimality within one second. The reason might be that when  $\theta_4 = 0$ , the problem is simplified to a location-allocation problem with rejection, whose computational complexity is much lower than our MCVAS problem. When  $\theta_4$  increases from one to five, the tardiness penalty cost decreases, and the average gap slightly increases. In general, the solutions are stable when tardiness penalty cost is considered, which means that the factor of tardiness has already been well considered in the original solution.

## 6. Case study

In this section, we conduct a case study based on the data from the mass COVID-19 vaccination program in Yiwu, a city in Zhejiang Province, China. Yiwu is making full efforts to improve the

**Table 9 Basic information of vaccination sites**

No.	Longitude	Latitude	$c_k$	Capacity	$Q$
1	120.1309	29.359996	500	3000	600
2	120.0598	29.323416	390		
3	120.083114	29.317479	320		
4	120.089429	29.292797	310		
5	120.176447	29.337946	420		

vaccination rate as quickly as possible. Many temporary vaccination sites have been built, and communities and companies can make appointments. We focus on five of the 14 subdistricts and choose 63 main communities to form the case. Note that the basic data is mainly found on the government website of Yiwu (Yiwu Municipal People’s Government), and we further generated some necessary parameters based on real situations to make the case applicable to our study.

### 6.1. Data Source

The locations of the vaccination sites and communities are visualized in Figure 2. Table 9 and 10 report the longitude and latitude of the vaccination sites and communities, along with some necessary parameters. All the vaccination sites are well-equipped with all the required facilities and equipment for the COVID-19 vaccination implementation and are operated by medical workers and volunteers. Therefore, the fixed opening cost  $c_k$  of each vaccination site  $k$  is set  $c_k \in \{300, 310, \dots, 500\}$  ( $\times 10^2$  RMB). The processing time of appointment  $j$  is computed using the number of recipients divided by the number of recipients that can be vaccinated per minute. With multiple vaccination stations launched in each vaccination site, the capacity of each site is about 3000 recipients per day, which means that they can vaccinate five recipients per minute on average. The number of recipients in each appointment, denoted as  $N_{rep}$ , ranges from the interval  $[100, 500]$ , thus  $p_j = N_{rep}/5$  and  $q_j = N_{rep}/100$ , both rounding to the nearest integer. Based on Basciftci et al. (2021) and Guo et al. (2021), the rejection cost is set to be 6 ( $\times 10^2$  RMB) per recipient, i.e.,  $r_j = 30 * p_j$ .

### 6.2. Results

LBBD solves the case with a time limit of 3600 seconds, and the result is illustrated in Figure 3. The figure shows that 62 out of 63 appointments are accepted and assigned to the vaccination sites. Only appointment No. 62 is rejected due to the working time capacity and will be arranged to another working day, incurring a rejection cost of 1980. The result shows a balance between cost-saving and time accuracy. Based on the schedule in Figure 3, we can inform the recipients of specific information, including the assigned vaccination site and the due arrival time.

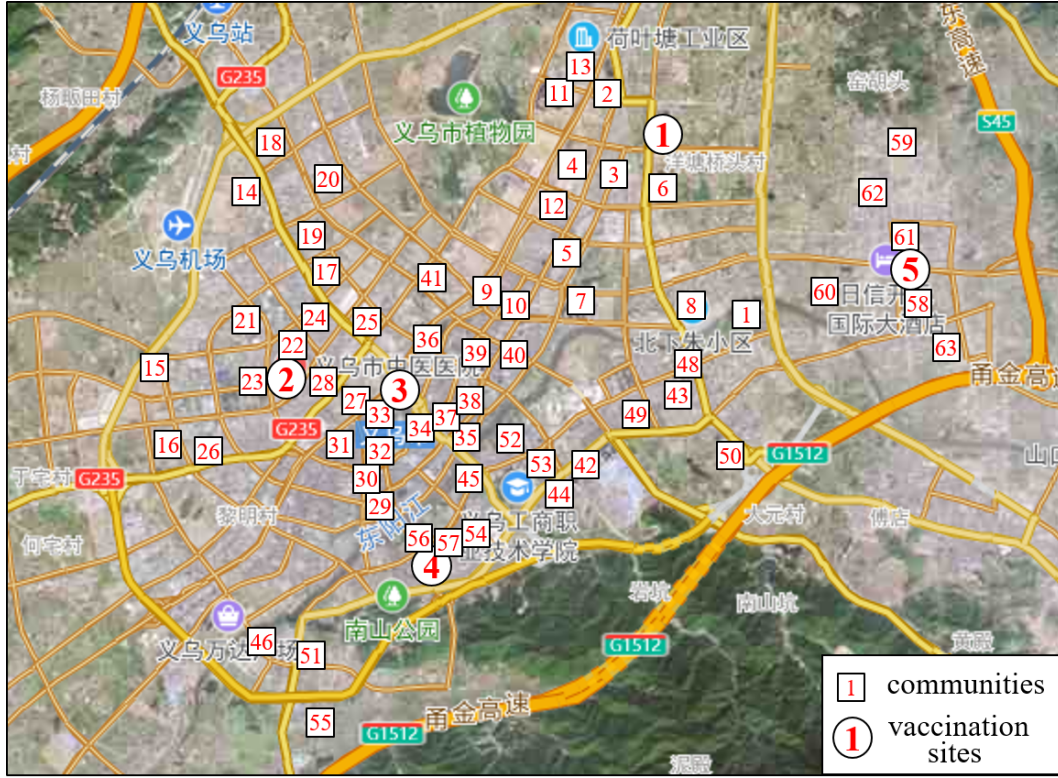


Figure 2 Locations of communities and vaccination sites

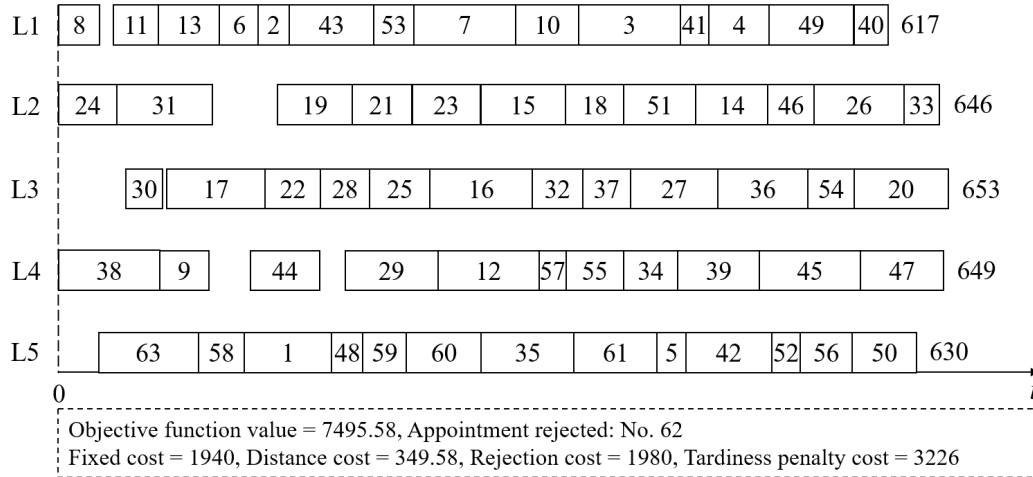


Figure 3 Result of the case study

## 7. Managerial implications

Our model provides an executable plan for decision-makers during the mass COVID-19 vaccination. An optimal solution of our model helps decision-makers by providing suggestions on: 1) which vaccination site to open; 2) which appointment to accept or reject; 3) which appointment is assigned to each opened vaccination site; and 4) the service time of each appointment at its assigned vacci-

**Table 10 Basic information of communities**

No.	Longitude	Latitude	$p_j$	$e_j$	$l_j$	No.	Longitude	Latitude	$p_j$	$e_j$	$l_j$
1	120.144174	29.338651	64	70	220	33	120.078602	29.313952	26	460	540
2	120.123052	29.368453	23	130	210	34	120.086661	29.312234	40	400	480
3	120.124009	29.354498	75	320	470	35	120.093797	29.310586	68	240	320
4	120.115728	29.356476	44	380	460	36	120.088409	29.32694	66	440	520
5	120.109493	29.343224	21	280	430	37	120.091995	29.314104	35	360	460
6	120.128893	29.35232	29	110	260	38	120.095416	29.316186	74	0	70
7	120.114935	29.336798	75	250	350	39	120.095173	29.325738	60	370	520
8	120.137664	29.333303	30	0	130	40	120.104255	29.325556	25	340	430
9	120.098654	29.334356	36	70	150	41	120.08708	29.337057	21	400	470
10	120.102587	29.332072	46	270	390	42	120.116269	29.305538	63	450	520
11	120.115061	29.368522	34	40	190	43	120.134751	29.319296	62	170	320
12	120.111206	29.349361	74	250	340	44	120.111573	29.304353	51	140	240
13	120.119539	29.377651	45	70	220	45	120.095265	29.30536	74	430	550
14	120.054922	29.348419	53	420	490	46	120.056229	29.278145	34	430	550
15	120.036689	29.321656	62	290	360	47	120.077538	29.27389	61	450	570
16	120.040614	29.307743	75	270	420	48	120.136689	29.322096	23	120	270
17	120.072154	29.334468	72	80	160	49	120.126944	29.31555	62	420	570
18	120.057826	29.360957	43	310	410	50	120.145746	29.309051	47	380	460
19	120.063177	29.345607	55	160	280	51	120.064853	29.275351	53	390	470
20	120.069617	29.35121	69	430	550	52	120.102939	29.310325	21	240	390
21	120.05154	29.328273	44	190	280	53	120.108698	29.308391	29	240	320
22	120.062172	29.326834	41	140	240	54	120.081039	29.290155	34	420	540
23	120.0562	29.320191	50	250	320	55	120.06776	29.265219	42	290	440
24	120.066757	29.332415	43	0	50	56	120.085405	29.292966	38	490	590
25	120.0769	29.32799	44	190	310	57	120.094257	29.293371	20	220	340
26	120.049016	29.308078	66	290	360	58	120.182548	29.333045	34	80	150
27	120.073904	29.316033	64	390	470	59	120.174238	29.357449	32	180	260
28	120.070172	29.320239	36	160	230	60	120.169032	29.335254	55	190	310
29	120.078043	29.300721	68	210	330	61	120.177667	29.344364	61	340	420
30	120.076118	29.301525	27	50	200	62	120.173353	29.351007	66	130	220
31	120.070974	29.31053	70	0	140	63	120.185392	29.324978	73	30	130
32	120.077813	29.311684	37	310	400						

nation site. The above decisions are made under a well-defined objective, which properly considers the operational cost for opening a vaccination site, the acceptance or rejection of an appointment, the travel distance of vaccine recipients, and the tardiness of vaccination service. Based on the previous sections' numerical results and sensitivity analysis, we next present some practical tips in establishing and operating the mass vaccination appointment scheduling program.

First, the ambition of the mass COVID-19 vaccination programs are to vaccinate as many recipients as possible, to assign the recipients to nearby vaccination stations, and to vaccinate the recipients within the time slot they appointed. The conflicts among the government's ambition, the recipients' demand, and the limited capacity call for a tradeoff between the operational cost and service level. The operational cost refers to the cost for opening a vaccination site. At the

same time, the service level is determined by minimizing the total travel distance, the appointment rejection cost, and the service tardiness cost. The four components to be optimized in the objective function allow decision-makers to obtain solutions that meet the requirements of all interests. Also, the sensitivity analysis shows that our model can suit the decision-makers' preferences by adjusting the weights of different cost components. For example, if decision-makers focus on the cost-effectiveness of the vaccination program, they tend to allocate a more significant weight to the cost of opening a vaccination site. In this case, the model would care more about the cost component, and an optimal solution may enable fewer vaccination sites to open. This may lead to a longer travel distance of vaccine recipients, a higher possibility of appointment rejection, and service tardiness. Thus, it is essential to consider the total travel distance, rejection cost, and tardiness in the objective function to balance the cost and service level, such that decision-makers can make flexible decisions by varying the weights of different components.

It is beneficial to optimize the total weighted travel distance of vaccine recipients. First, the model tends to open vaccination sites close to vaccine recipients. Second, a short travel distance of recipients reduces the potential risk of cross-infection during epidemics. Third, a short travel distance of recipients improves the service level by providing an excellent vaccinating experience. However, always allocating recipients to their nearest vaccination site may lead to a long waiting time (tardiness) due to the uneven distribution of the vaccination population.

To avoid a long wait and work overload, we introduce two cost components in the objective, i.e., to minimize the rejection cost component and the total weighted tardiness component. We expect to accept as many appointments as possible to finish the vaccination program. However, due to the limited vaccine supply and vaccination capacity at each site, it is impractical to accept all appointments. In this case, the model must determine which appointment to accept and which to reject and reassign to other dates and arrange the service sequences based on the vaccine recipients' time windows. These considerations may improve the service level of the vaccination program.

The vaccination appointment scheduling problem is a complicated combinatorial optimization problem. The developed LBBD finds optimal or near-optimal solutions for instances with up to 100 appointments in acceptable computation time. The proposed matheuristic MH is applicable to solve practical-sized instances with hundreds of appointments and dozens of vaccination sites. Therefore, our work provides a systematic solution for mass vaccination programs with the developed model and algorithms.

## 8. Conclusions

This study investigated a mass vaccination problem during pandemics or epidemics. Given the limited capacity of vaccine supply, certain personnel such as healthcare workers and education

faculties have high priority in vaccination. We identify an appointment scheduling problem in vaccinating a large population, denoted as the MCVAS problem. The problem jointly optimizes the vaccination selection decisions, appointment selection decisions, appointment assignment decisions, and appointment sequencing decisions. The objective is to minimize the weighted sum of the total cost for operating vaccination sites, total travel distance of vaccine recipients, rejection cost, and penalty cost incurred by service tardiness. We formally describe and formulate the problem into a MILP. We then develop an exact logic-based Benders decomposition (LBBD) method for the problem. The LBBD method decomposes the MCVAS problem into a master problem (MP) concerning vaccination site selection, appointment acceptance, appointment assignment decisions, and a subproblem (SP) dealing with the sequencing of appointments at each vaccination site. The MP is then solved using a B&C method with the *branch and check* implementation. During the branching process of the B&C, Benders cuts are iteratively generated and added to MP upon finding a feasible solution to the MP. The branching process continues with added Benders cuts until the optimal solution is identified. We further develop a matheuristic method (MH) based on predetermined sequences to tackle practical-sized instances. The MH first defines a new set that contains all appointments sorted in a given rule. Then an approximate MILP model AP is formulated using the newly-defined set. The AP model is solved using an off-the-shelf solver to provide near-optimal solutions, which are further improved by solving a series of single machine scheduling problems with release date and tardiness minimization. We conducted numerical experiments on instances with up to 500 appointments and 50 vaccination sites. In instances with up to 100 appointments and 20 vaccination sites, the LBBD method generally outperforms LO and MH in providing more optimal solutions. The LBBD method and MH are superior to the LO model in providing a smaller relative optimality gap in a shorter computation time. In particular, MH is time-efficient in solving these instances in general. Results on practical-sized instances with up to 500 appointments and 20 vaccination sites indicate that MH and LBBD are superior to LO with lower average gaps. Primarily, MH obtains an average gap of 5.64% on these instances. We conducted a sensitivity analysis to evaluate the effects of the four cost components. The result provides valuable managerial insights for practitioners. We further draw managerial implications from the result of the numerical experiments.

The mass vaccination appointment scheduling problem is an interesting topic that deserves further study. We indicate a few research perspectives. First, the problem can be extended to cover multiple periods, such that a more integrated plan can be achieved. In this case, the problem becomes more complex, and efficient algorithms should be developed to handle practical-sized instances. Second, the service duration of each appointment is often uncertain. Thus, it is essential to generalize the problem to consider service time uncertainty. One possible method is to formulate



the problem as a two-stage stochastic program. The first stage determines the vaccination site selection, the appointment selection, and the appointment assignment without exactly knowing the service duration. The second stage sequences the appointments with the realization of the random service duration. Then the sample average approximation or L-shaped methods may be applied to solve the stochastic counterpart of the MCVAS problem.

## References

- S. Akbarinasaji and A. R. Mckendall. Heuristics for the integrated single machine scheduling and location problem. *International Journal of Industrial and Systems Engineering*, 27(2):196–209, 2017.
- V. V. Albani, J. Loria, E. Massad, and J. P. Zubelli. The impact of COVID-19 vaccination delay: A data-driven modeling analysis for Chicago and New York City. *Vaccine*, 39(41):6088–6094, 2021.
- R. Alizadeh Foroutan, J. Zeidi, M. Abedi, and R. Chiong. A modified genetic algorithm for non-emergency outpatient appointment scheduling with highly demanded medical services considering patient priorities. *Computers & Industrial Engineering*, 139:106106, 2020.
- Arizona Department of Health Services. Arizona COVID-19 vaccine prioritization. <https://azdhs.gov/documents/preparedness/epidemiology-disease-control/infectious-disease-epidemiology/novel-coronavirus/vaccine/covid19-vp-infographic-eng.pdf> Accessed: October 26, 2021.
- Y. Bartal, S. Leonardi, A. Marchetti-Spaccamela, J. Sgall, and L. Stougie. Multiprocessor scheduling with rejection. *SIAM Journal on Discrete Mathematics*, 13(1):64–78, 2000.
- B. Basciftci, X. Yu, and S. Shen. Resource distribution under spatiotemporal uncertainty of disease spread: Stochastic versus robust approaches. available at <http://dx.doi.org/10.2139/ssrn.3799367>, 2021.
- G. Bektur and T. Sara. A mathematical model and heuristic algorithms for an unrelated parallel machine scheduling problem with sequence-dependent setup times, machine eligibility restrictions and a common server. *Computers & Operations Research*, 103:46–63, 2019.
- I. Bendavid and B. Golany. Predetermined intervals for start times of activities in the stochastic project scheduling problem. *Annals of Operations Research*, 186(1):429–442, 2011.
- J. F. Benders. Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 1962.
- P. Bennett, J. Crosbie, and P. Dick. Use of or by government to inform health policy in England: Examples and reflections. *Operations Research for Health Care*, 1(1):1–5, 2012.
- K. Bülbül and H. Şen. An exact extended formulation for the unrelated parallel machine total weighted completion time problem. *Journal of Scheduling*, 20(4):373–389, 2016.
- Centers for Disease Control and Prevention. COVID data tracker. <https://covid.cdc.gov/covid-data-tracker/#vaccinations> Accessed: October 26, 2021.

- T. C. E. Cheng, C. T. Ng, J. J. Yuan, and Z. H. Liu. Single machine scheduling to minimize total weighted tardiness. *European Journal of Operational Research*, 165(2):423–443, 2005.
- R. Cordone and P. Hosteins. A bi-objective model for the single-machine scheduling problem with rejection cost and total tardiness minimization. *Computers & Operations Research*, 102:130–140, 2019.
- T. Dai. How to fix the mess of COVID-19 vaccine appointment scheduling. <https://www.fastcompany.com/90607146/how-to-fix-the-mess-of-covid-19-vaccine-appointment-scheduling>, 2021.
- Q. V. Dang, T. V. Diessen, T. Martagan, and I. Adan. A matheuristic for parallel machine scheduling with tool replacements. *European Journal of Operational Research*, 291(2):640–660, 2020.
- S. De Vuyst, H. Bruneel, and D. Fiems. Computationally efficient evaluation of appointment schedules in health care. *European Journal of Operational Research*, 237(3):1142–1154, 2014.
- G. Dósa and Y. He. Scheduling with machine cost and rejection. *Journal of Combinatorial Optimization*, 12(4):337–350, 2006.
- M. E. Dyer and L. A. Wolsey. Formulating the single machine sequencing problem with release dates as a mixed integer program. *Discrete Applied Mathematics*, 26(2-3):255–270, 1990.
- M. Elhafsi. Optimal leadtimes planning in serial production systems with earliness and tardiness costs. *Iie Transactions*, 34(3):233–243, 2002.
- S. E. Elmaghraby, A. A. Ferreira, and L. V. Tavares. Optimal start times under stochastic activity durations. *International Journal of Production Economics*, 64(1-3):153–164, 2005.
- D. Elvikis, H. W. Hamacher, and M. T. Kalsch. Simultaneous scheduling and location (Scheloc): the planar ScheLoc makespan problem. *Journal of Scheduling*, 12(4):361–374, 2009.
- L. Fanjul-Peyro, F. Perea, and R. Ruiz. Models and matheuristics for the unrelated parallel machine scheduling problem with additional resources. *European Journal of Operational Research*, 260(2):482–493, 2017.
- J. Feldman, N. Liu, H. Topaloglu, and S. Ziya. Appointment scheduling under patient preference and no-show behavior. *Operations Research*, 62(4):794–811, 2014.
- B. H. Foy, B. Wahl, K. Mehta, A. Shet, G. I. Menon, and C. Britto. Comparing COVID-19 vaccine allocation strategies in India: A mathematical modelling study. *International Journal of Infectious Diseases*, 103:431–438, 2021.
- P. M. França, A. Mendes, and P. Moscato. A memetic algorithm for the total tardiness single machine scheduling problem. *European Journal of Operational Research*, 132(1):224–242, 2001.
- G. P. Georgiadis and M. C. Georgiadis. Optimal planning of the COVID-19 vaccine supply chain. *Vaccine*, 39(37):5302–5312, 2021.
- R. Graham, E. Lawler, J. Lenstra, and A. Kan. Optimization and approximation in deterministic sequencing and scheduling: a survey. In P. Hammer, E. Johnson, and B. Korte, editors, *Discrete Optimization II*, volume 5 of *Annals of Discrete Mathematics*, pages 287–326. Elsevier, 1979.

- 
- F. Grenouilleau, N. Lahrichi, and L.-M. Rousseau. New decomposition methods for home care scheduling with predefined visits. *Computers & Operations Research*, 115:104855, 2020.
- C. Guo, M. Bodur, D. M. Aleman, and D. R. Urbach. Logic-based Benders decomposition and binary decision diagram based approaches for stochastic distributed operating room scheduling, 2021.
- H. Hennes and H. W. Hamacher. *Integrated scheduling and location models: single machine makespan problems*. Technical Report, University of Kaiserslautern, Shaker Verlag, Aachen, 2002.
- C. Heßler and K. Deghdak. Discrete parallel machine makespan ScheLoc problem. *Journal of Combinatorial Optimization*, 34(4):1159–1186, 2017.
- J. N. Hooker and G. Ottosson. Logic-based Benders decomposition. *Mathematical Programming*, 96(1):33–60, 2003.
- S. Hovav and D. Tsadikovich. A network flow model for inventory management and distribution of influenza vaccines through a healthcare supply chain. *Operations Research for Health Care*, 5:49–62, 2015.
- N. Huynh. Reducing truck turn times at marine terminals with appointment scheduling. *Transportation Research Record*, (2100):47–57, 12 2009.
- M. Ji, J. Wang, and W. Lee. Minimizing resource consumption on uniform parallel machines with a bound on makespan. *Computers & Operations Research*, 40(12):2970–2974, 2013.
- J. Jiao, Z. Liu, and S. Cai. Dynamics of an SEIR model with infectivity in incubation period and homestead-isolation on the susceptible. *Applied Mathematics Letters*, 107:106442, 2020.
- M. T. Kalsch and Z. Drezner. Solving scheduling and location problems in the plane simultaneously. *Computers & Operations Research*, 37(2):256–264, 2010.
- M. R. Korupolu, C. Plaxton, and R. Rajaraman. Analysis of a local search heuristic for facility location problems. *Journal of Algorithms*, 37(1):146–188, 2000.
- R. Lahyani, K. Chebil, M. Khemakhem, and L. C. Coelho. Matheuristics for solving the multiple knapsack problem with setup. *Computers & Industrial Engineering*, 129:76–89, 2019.
- K. Li, Y. Shi, S. lin Yang, and B. yi Cheng. Parallel machine scheduling problem to minimize the makespan with resource dependent processing times. *Applied Soft Computing*, 11:5551–5557, 12 2011.
- X. Li, Y. Pan, S. Jiang, Q. Huang, Z. Chen, M. Zhang, and Z. Zhang. Locate vaccination stations considering travel distance, operational cost, and work schedule. *Omega*, page 102236, 2020.
- J. Lim, E. Claypool, B. A. Norman, and J. Rajgopal. Coverage models to determine outreach vaccination center locations in low and middle income countries. *Operations Research for Health Care*, 9:40–48, 2016.
- M. Liu, X. Liu, E. Zhang, F. Chu, and C. Chu. Scenario-based heuristic to two-stage stochastic program for the parallel machine ScheLoc problem. *International Journal of Production Research*, 57(6):1706–1723, 2019.

- V. Maniezzo, T. Stützle, and S. Voss. *Matheuristics – Hybridizing Metaheuristics and Mathematical Programming*. Springer Publishing Company, Inc., 2010.
- Milken Institute and FirstPerson. COVID-19 vaccine tracker. <https://www.covid-19vaccinetracker.org/> Accessed: October 26, 2021.
- B. Naderi and V. Roshanaei. Branch-Relax-and-Check: A tractable decomposition method for order acceptance and identical parallel machine scheduling. *European Journal of Operational Research*, 286(3): 811–827, 2020.
- B. Naderi, K. Govindan, and H. Soleimani. A Benders decomposition approach for a real case supply chain network design with capacity acquisition and transporter planning: wheat distribution network. *Annals of Operations Research*, 291:685–705, 2020.
- A. Nagurney. Optimization of supply chain networks with inclusion of labor: Applications to COVID-19 pandemic disruptions. *International Journal of Production Economics*, 235:108080, 2021.
- J. Ou, X. Zhong, and G. Wang. An improved heuristic for parallel machine scheduling with rejection. *European Journal of Operational Research*, 241(3):653–661, 2015.
- Our World in Data. Coronavirus (COVID-19) Vaccinations. <https://ourworldindata.org/covid-vaccinations> Accessed: October 26, 2021.
- M. Pinedo and E. Rammouz. A note on stochastic scheduling on a single machine subject to breakdown and repair. *Probability in the Engineering Information Sciences*, 2(01):41–49, 1988.
- V. Roshanaei, C. Luong, D. M. Aleman, and D. R. Urbach. Reformulation, linearization, and decomposition techniques for balanced distributed operating room scheduling. *Omega*, 93:102043, 2020.
- H. Şen and K. Bülbül. A strong preemptive relaxation for weighted tardiness and earliness/tardiness problems on unrelated parallel machines. *INFORMS Journal on Computing*, 27(1):135–150, 2015.
- S. O. Shim and Y. D. Kim. Scheduling on parallel identical machines to minimize total tardiness. *European Journal of Operational Research*, 177(1):135–146, 2007.
- B. Shnits, I. Bendavid, and Y. N. Marmor. An appointment scheduling policy for healthcare systems with parallel servers and pre-determined quality of service. *Omega*, 97:102095, 2020.
- M. Soltani, M. Samorani, and B. Kolfal. Appointment scheduling with multiple providers and stochastic service times. *European Journal of Operational Research*, 277(2):667–683, 2019.
- D. Sun, L. Tang, and R. Baldacci. A Benders decomposition-based framework for solving quay crane scheduling problems. *European Journal of Operational Research*, 273(2):504–515, 2019.
- Suzhou Municipal People’s Government. <http://www.suzhou.gov.cn/> Accessed: November 14, 2021.
- Q. C. Ta, J.-C. Billaut, and J.-L. Bouquard. Matheuristic algorithms for minimizing total tardiness in the m-machine flow-shop scheduling problem. *Journal of Intelligent Manufacturing*, 29:617–628, 2015.

- 
- S. Tanaka and M. Araki. An exact algorithm for the single-machine total weighted tardiness problem with sequence-dependent setup times. *Computers & Operations Research*, 40(1):344–352, 2013.
- S. Tanaka and S. Fujikuma. A dynamic-programming-based exact algorithm for general single-machine scheduling with machine idle time. *Journal of Scheduling*, 15(3):347–361, 2012.
- S. Tanaka and S. Sato. An exact algorithm for the precedence-constrained single-machine scheduling problem. *European Journal of Operational Research*, 229(2):345–352, 2013.
- E. Vallada and R. Ruiz. A genetic algorithm for the unrelated parallel machine scheduling problem with sequence dependent setup times. *European Journal of Operational Research*, 211(3):612–622, 2011.
- J. Wang, Y. Peng, H. Xu, Z. Cui, and R. Williams. The COVID-19 vaccine race: Challenges and opportunities in vaccine formulation. *AAPS PharmSciTech*, 21(6), 2020a.
- K. Wang, Y. Shao, and W. Zhou. Matheuristic for a two-echelon capacitated vehicle routing problem with environmental considerations in city logistics service. *Transportation Research Part D: Transport and Environment*, 57:262–276, 2017.
- S. Wang, R. Wu, F. Chu, J. Yu, and X. Liu. An improved formulation and efficient heuristics for the discrete parallel-machine makespan ScheLoc problem. *Computers & Industrial Engineering*, 140(Feb.): 106238.1–106238.9, 2020b.
- Worldometer. COVID-19 coronavirus pandemic. <https://www.worldometers.info/coronavirus/> Accessed: October 26, 2021.
- Yiwu Municipal People’s Government. <http://www.yw.gov.cn/> Accessed: November 14, 2021.
- L. Zhang, L. Lu, and J. Yuan. Single machine scheduling with release dates and rejection. *Computers & Operations Research*, 198(3):975–978, 2009.
- Y. Zhang, W.-H. Lin, M. Huang, and X. Hu. Multi-warehouse package consolidation for split orders in online retailing. *European Journal of Operational Research*, 289(3):1040–1055, 2021.
- J. Zhao, B. Wu, and G. Y. Ke. A bi-objective robust optimization approach for the management of infectious wastes with demand uncertainty during a pandemic. *Journal of Cleaner Production*, 314:127922, 2021.
- X. Zhong and J. Ou. Parallel machine scheduling with restricted job rejection. *Theoretical Computer Science*, 690:1–11, 2017.
- S. Zhou, D. Li, and Y. Yin. Coordinated appointment scheduling with multiple providers and patient-and-physician matching cost in specialty care. *Omega*, page 102285, 2020.