Simulation optimization in healthcare resource planning: A literature review

Lien Wang [a] and Erik Demeulemeester [a]
[a] KU Leuven, Faculty of Business and Economics, Department of Decision Sciences and Information Management, Research Centre for Operations Management, Naamsestraat 69, 3000 Leuven, Belgium

Abstract

In healthcare, the planning and the management of resources are challenging as there are always a number of complex and stochastic factors in both demand and supply. Simulation optimization (SO) that combines simulation analysis and optimization techniques is well suited for solving complicated, stochastic, and mathematically intractable decision problems. In order to comprehensively unveil the degree to which SO has been used to solve healthcare resource planning problems, this paper reviews the academic articles published until 2021 and categorizes them into multiple classification fields that are related to either problem perspectives (i.e., healthcare services, planning decisions, and objectives) or methodology perspectives (i.e., SO approaches and applications). We also examine the relations between the individual fields. We find that emergency care services are the most applied domain of SO, and that discrete-event simulation and random search methods (especially genetic algorithms) are the most frequently used methods. The literature classification can help researchers quickly learn this research area and identify the publications of interest. Finally, we identify major trends, insights and conclusions that deserve special attention when studying this area. We suggest many avenues for further research that provide opportunities for expanding existing methodologies and for narrowing the gap between theory and practice.

Keywords: Health care management; Simulation optimization; Operations management; Resource capacity planning; Literature review.

1 Introduction

Healthcare providers face a continuous challenge to plan and control their healthcare delivery processes efficiently since the demand for healthcare services and the related expenditures are growing [124]. Within a healthcare organization, planning and control involve defining goals and deciding on what to do, how to do it, when to do it, and who should do it [73]. The main objective of healthcare planning and control is to provide high-quality and highly efficient services at the lowest possible cost to society. It consists of multiple managerial areas, e.g., making medical, resource capacity, and financial decisions. In this review, we focus on the managerial area of resource planning in healthcare. Specifically, this area makes decisions on the dimensioning, planning, scheduling, and control of various renewable healthcare resources, e.g., operating rooms, ward beds, intensive care units (ICUs), and staff [66].

However, making these decisions is not straightforward as the real-world healthcare environment is quite complex, dynamic, and stochastic, which often results in an analytically intractable problem. In addition, the healthcare system often involves many different stakeholders, such as patients, medical staff, employers, and government, who might have conflicting preferences and goals [133]. The Operations Research and Management Sciences (OR/MS) community has taken on these challenges for many years by developing various methods to improve the healthcare system efficiency, e.g., computer simulation, mathematical optimization, probability and stochastic models, and decision analysis [75].

Among these methods, simulation optimization (SO) is a potentially useful and flexible approach for solving complicated, stochastic, and mathematically intractable decision problems, without having to make many restrictive assumptions [125]. Simulation optimization (also known as optimization via simulation or simulation-based optimization) refers to the optimization of a given objective function satisfying some constraints, both of which can be estimated through a stochastic simulation [53, 5]. It searches for optimal input settings to the stochastic simulation such that the target objective is optimized. In contrast to algebraic model-based mathematical programming, the simulation in SO serves as a black box to model complex, dynamic, and stochastic systems. In recent years, advancements in computer computing power have increased the popularity of SO methods in diverse fields such as manufacturing, transportation, medicine and biology [5, 108].

A general simulation optimization problem can be expressed by the following [5]:

$$\min \ \mathbb{E}_\omega[f(x,y,\omega)] \tag{1}$$

$$\text{s.t. } \mathbb{E}_\omega[g(x,y,\omega)] \leq 0 \tag{2}$$

$$h(x,y) \leq 0 \tag{3}$$

$$x_l \leq x \leq x_u \tag{4}$$

$$x \in \mathbb{R}^n, y \in \mathbb{D}^m. \tag{5}$$

The SO problem is to minimize an objective function, which has to be estimated by simulation. The objective function takes the form of an expected value (i.e., $\mathbb{E}_\omega[f(x, y, \omega)]$) for a specific instance of continuous input variables x and discrete input variables y, with a realization of random variables in the simulation, i.e., the vector ω (which might or might not be a function of x and y). Similarly, the stochastic constraints expressed by the vector-valued function g that involves random variables are also evaluated by the simulation and expected values are used. There might be other constraints (defined by $h(x, y) \leq 0$) that do not include random variables, and bound constraints on the input variables. In addition, the objective function can take other forms of measures, e.g., quantiles or the variance.

This paper reviews the articles involving the application of SO in healthcare resource planning. The purpose of this review paper is threefold. First, we aim to unveil the degree to which SO has been used to solve the healthcare resource planning problem and to identify any trends that have arisen in this research field. Second, we want to structure the published literature in a way that recent research contributions can easily be compared on multiple angles and connected to each other. Furthermore, interesting patterns might be revealed and opportunities for future work can be identified. Third, the resulting classification can help researchers to quickly learn about this field and to quickly identify the papers that are related to their needs. In addition, we discuss the situations in which SO is most useful, the complementary relation between SO and other OR/MS methods, the key aspects to consider and the recent advances in SO that can potentially be applied to healthcare planning.

2 Literature review methodology

2.1 *Literature collection, identification and analysis*

We used a structured literature search method to ensure that we found key and state-of-the-art contributions that are under the scope of this review. Table 1 provides a schematic overview of our search method. First, we identified search terms and used wildcards in the search terms. These terms include methodology-related terms and healthcare-related terms. As for the former, "simulation" and "optimization" are used. These two notions can be combined in various ways: optimization with simulation-based iterations, simulation with optimization-based iterations, sequential simulation and optimization, and alternate simulation and optimization [51]. We focus on the papers related to the first category, i.e., where the optimization problem has to work with evaluations of the stochastic objective function $f(x, y, \omega)$.

Table 1. The literature search method.

Step	What to do
Step 1:	Identify search terms related to the methodology and to healthcare
Step 2:	Search the subject categories of OR&MS and HCS&S in WoS
Step 3:	Select a base set: the 20 most-cited articles relevant for our review
Step 4:	Perform a backward and forward search on the base set articles
Step 5:	Include all articles from WoS core collection journals
Step 6:	Search in other databases

In terms of the healthcare-related terms, only using the general terms, such as healthcare, hospital, and patient, is not sufficient, since the healthcare delivery involves many different care services provided by various healthcare organizations. According to the taxonomic classification of healthcare resource planning decisions in Hulshof et al. [73], we consider six care services (i.e., ambulatory care services (ACS), emergency care services (ECS), surgical care services (SCS), inpatient care services (ICS), residential care services (RCS) and home care services (HCS)) in order to obtain a more comprehensive pool of articles. Briefly, ACS provide care for patients without an overnight stay in the hospital, e.g., outpatient clinics, primary care services, and hospital radiology departments. ECS face urgent and emergent medical problems, e.g., hospital emergency departments, ambulances, and trauma centers. SCS perform operative procedures (surgeries) on patients, e.g., operating theaters and surgical day care centers. ICS care for hospitalized patients, e.g., regular wards, intensive care units, and neonatal care units. RCS provide long-term care and assistance for frail elderly populations, e.g., nursing homes. HCS are provided for patients by visiting medical staff at the patients' homes, e.g., medical care at home.

Since Hulshof et al. [73] have identified an extensive list of search terms for each care service, we use similar search terms (see Table 2) to them. Using those search terms, we performed an initial search for academic papers published until August 2021 based on the database of Web of Science (WoS). WoS was chosen as it provides the possibility to select Operations Research & Management Science (OR&MS) and Health Care Sciences & Services (HCS&S) as specific subject categories (both are closely related to the topic of this review) and to sort search results on the number of citations. Refer to the WoS Core

Collection website (i.e., https://mjl.clarivate.com/home) for a full list of journals within OR&MS and HCS&S. The search fields involve titles, abstracts, or keywords. Only peer-reviewed journal articles and articles written in English are considered in this review.

Table 2. Details of literature search terms.

Services	Care term	Methodology term
General	healthcare ∪ health care ∪ medical ∪ hospital ∪ patient	simulation ∩ optimization
ACS	outpatient ∪ ambulatory ∪ appointment ∪ clinic ∪ primary care ∪ radiology ∪ diagnostic service$ ∪ diagnostic facilit* ∪ general practi* ∪ community service$	
ECS	emergenc* ∪ emergent ∪ acute ∪ urgenc* ∪ urgent ∪ accident ∪ ambulance ∪ trauma	
SCS	operating room$ ∪ operating theat* ∪ operating suite ∪ surger* ∪ surgical	
ICS	inpatient ∪ bed$ ∪ ward$ ∪ intensive care ∪ neonatal care	
RCS	nursing home$ ∪ mental care ∪ psychiatric ∪ rehabilitation ∪ long-term care ∪ retirement ∪ elderly ∪ geriatric	
HCS	home care ∪ home health* home-care ∪ home-health* ∪ care at home	

Notes. A search engine can replace $ by zero or one character. A search engine can replace '*' by any group of characters, including no character.

After obtaining an initial list of articles, we expanded this list by performing a backward and forward search. This search was based on a base set containing the 20 most-cited articles that are selected from the initial list. Specifically, we identified all articles that are referred by or refer to one of the articles in the base set and that deal with the application of SO to healthcare planning. We included relevant articles from all WoS Core Collection journals at this forward and backward search stage. Furthermore, we also performed a search in the databases of PubMed and Scopus with our search terms. This search did not produce significant additions to the already identified set of articles. The resulting articles were then carefully examined to see whether they are under scope for this review. Finally, 141 papers are selected for inclusion, which were published between 1973 and 2021.

Figure 1 clearly shows that the application of SO in healthcare resource planning is becoming increasingly popular in the literature. Most of the papers were published in recent years, especially from 2013 onwards, although SO has had a long and well-known history of developments [54]. The reason might be, on the one hand, that the ever-improving healthcare information systems worldwide in recent years [121] provide ample availability of data to allow for simulation modeling and optimization. On the other hand, with the

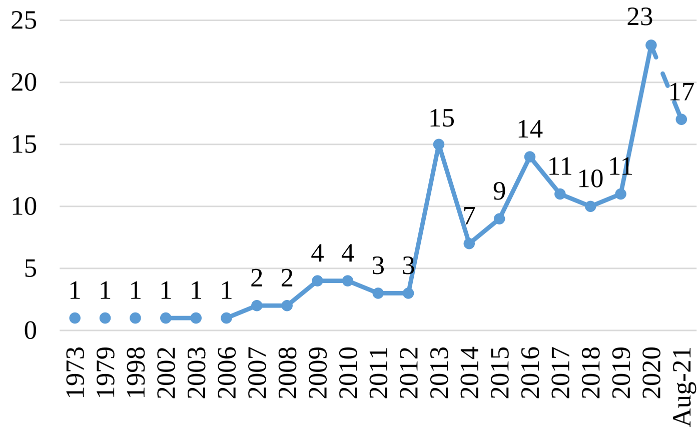

Figure 1. The number of published papers per year.

fast development in computer computing powers, SO is expected to increase its impact on healthcare problems. There are a total of 60 peer-reviewed journals publishing the selected papers. Figure 2 shows the journals with at least 3 publications. As can be observed, Health Care Management Science is the leading journal on this topic with 13 papers, followed by the European Journal of Operational Research (11 papers).

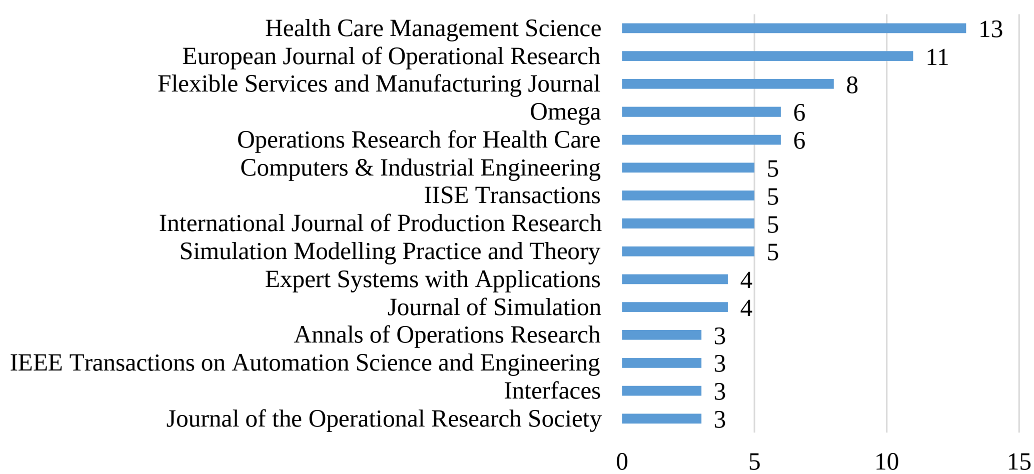

Figure 2. The number of published papers per peer-reviewed journal.

2.2 *Organization of the review*

We propose to use the classification fields in the order of how they should be tackled by authors when solving the healthcare resource planning problems of interest, which can well serve the purposes of this review. These classification fields are shown as follows:

- *What is the scope*: reviewing the literature according to the various healthcare services and the key decisions that need to be made, along with a discussion on the main problem characteristics of the application.

- *What are the objectives*: describing the objectives (e.g., access time, utilization, overtime, and financial value) and the type of objective functions (i.e., single/multiple objectives).
- *What are the SO approaches*: discussing the simulation modeling methods and the optimization algorithms/packages used in the SO modeling process, along with the type of decision variables (i.e., discrete or continuous) and the stochastic constraints.
- *Applying and validating the model*: indicating the information on the validation of simulation modeling, the testing (data) of the SO models, and the implementation in practice.

Each section includes a discussion of the specific field based on the selected papers and an explanation of the terminology when needed, along with a classification table that lists all relevant papers. Pooling these tables together serves as a reference tool to identify the subset of papers according to a certain characteristic. They additionally should enable the reader to learn how to solve the healthcare planning problem by the SO method. We also look into the relations between the individual classification fields by considering multiple dimensions simultaneously. Some sections discuss potential opportunities for future research.

3 What is the scope

When tackling the healthcare resource planning problem, the initial set of decisions to be made relates to the problem setting (e.g., care services and problem characteristics) and to the planning decisions. As for the decisions, researchers frequently distinguish between the strategic (long-term), the tactical (medium-term), the operational offline (short-term), and the operational online (real-time) decision levels. Table 3 classifies the literature based on the six care services and on the hierarchical decision levels.

3.1 *Strategic decision level*

The strategic decision level addresses the design, dimensioning and development of the healthcare delivery process based on aggregated information and forecasted demand. Overall, this decision level most frequently applies the SO method.

Table 3 shows that most SO studies that focus on this decision level are for emergency care services, relating to capacity dimensioning (e.g., determining the optimal resource

Table 3. Classification based on care services and planning decisions.

	Strategic	Tactical	Offline	Online
ACS	[90, 59, 38, 157]	[101, 86, 106, 32, 27, 131, 85, 84, 83, 82, 14, 167, 117, 72, 145, 138, 104, 168, 137, 159]	[129, 136, 159, 48, 49, 182, 40, 44, 20, 142, 155, 25, 8, 60]	[129]
ECS	[119, 162, 161, 115, 170, 26, 184, 116, 156, 177, 78, 37, 76, 56, 6, 61, 150, 98, 62, 127, 55, 178, 3, 35, 99, 57, 42, 34, 50, 58, 97, 112, 100, 77, 65, 135, 175, 149, 171]	[152, 151, 64, 173, 122, 144, 36]		[115, 26, 176, 184, 130, 1, 56, 4]
SCS	[107, 128, 123, 12]	[174, 164]	[93, 92, 43, 160, 9, 141, 154, 94, 118, 47, 169, 165, 105, 46, 183]	
ICS	[111, 71, 67, 110, 88, 185, 2, 166, 132, 126]	[179, 95]	[114, 113, 11, 163]	
RCS	[181, 180, 13]			
HCS			[148, 18]	

capacity in the hospital emergency department [3, 58, 35, 34]) and ambulance location (e.g., assigning available ambulances to different locations [116, 26, 76, 37]) decisions. This is due to the fact that the evaluation of these decisions usually needs to model highly complex and uncertain operational processes in which unscheduled patients with a higher urgency and with various degrees of illness or injury have to be provided with timely emergency medical services. This leads simulation to be a useful tool. For example, in the emergency departments, it is important to model patient flows from arrival to discharge, and to account for stochastic activities (e.g., patient arrivals and their associated care) [55, 42, 34, 97]. Furthermore, capacity dimensioning decisions often consider multiple resources (e.g., receptionists, doctors, nurses and beds) in the care activities, which will lead to a large number of combinations of decision variables (e.g., 4320 combinations in [3]), while the ambulance location problem over a network of bases has an exponentially large solution space [78, 115].

Other care services sometimes apply SO to handle capacity dimensioning decisions in order to best balance provider capacity and patient demand. Examples include determining the number of beds [166, 185, 132, 111] and staff [88, 132, 111] in the inpatient healthcare facilities (e.g., wards and ICUs), determining the number of operating rooms [128] and staff [107] in surgical care facilities. Most of these papers incorporate complex real-world characteristics, including random patient arrivals [88, 166, 185], uncertain service times (e.g., lengths of stay [88, 166, 185], and surgical procedure durations [107]), interactions among service activities [107, 132], and possible rejections [88, 166], etc.

3.2 *Tactical decision level*

At the tactical decision level, resource capacity settled at the strategic level is subdivided over different patient groups (e.g., disease type and urgency) and is spread over different time slots (e.g., dates and time intervals). Capacity levels can be temporarily adjusted between patient groups or on a day (e.g., overtime). Decisions at this level often demonstrate a combinatorial nature, e.g., allocating capacity to patient groups or time slots [64].

Most of the reviewed studies that focus on this decision level are for ambulatory care services, including decisions on capacity allocation, staff-shift scheduling (assigning staff to shifts), admission control (selecting patients from the waiting lists to be admitted) and appointment scheduling (determining the length of the appointment interval, etc). SO is typically chosen since it can search for high-quality solutions in a large search domain while simultaneously accounting for complex and stochastic factors in ambulatory facilities [85, 86, 131, 83]. Examples include uncertain patient demand [104, 168, 159, 86], variable service times [145, 168, 83, 86], unpunctuality [145, 85], no-shows/abandonment [131, 137, 83], walk-in patients [27, 167], and variable-interval rules [86, 85]. Furthermore, SO has an ability to incorporate complex probability distributions (e.g., an uncommon service time distribution [83, 159, 145, 86] and nonstationary patient arrivals [168, 27, 137]).

In emergency care services, Sinreich et al. [152] and Sinreich and Jabali [151] use simulation-based models to solve staff-scheduling problems as the complex patients' clinical pathways and the high degree of interactions between different staff members complicate such problems. In inpatient care services, some authors use the SO framework to analyze patient overflow strategies (to secondary inpatient units) [179, 95] and bed reallocation (among wards) [95] in the context of uncertain patient arrivals and lengths of stay (LOS).

3.3 *Operational decision level (offline and online)*

The operational level (both 'offline' and 'online') involves short-term decisions at the individual patient and resource levels. Offline operational planning is the planning of known patient demand in advance given fixed capacity, while online operational planning deals with the monitoring and control of the healthcare delivery process reacting to realized and unforeseen events. The operational decision level usually demonstrates an exponential growth of the solution space with any increase in the number of patients, in the time blocks of resources as well as in the random realizations of uncertain variables [160, 182].

In Table 3, two common examples of offline operational planning with SO are patient-to-appointment scheduling and surgical case scheduling. In these cases, detailed operational factors are often incorporated, e.g., uncertain arrivals of unplanned patients [49, 160], patient unpunctuality [129, 44, 40, 160], stochastic service durations [129, 136, 44, 160, 43], no-shows [129, 136, 44, 40], and heterogeneous sequences of services [142]. Each patient/surgery type is often assumed to have its own probability distribution function for the uncertain activities [44, 142, 105, 43]. Patient behavioral factors (e.g., selection and abandonment [49]) and server behavioral factors (i.e., a service speedup in response to congestion [182]) are sometimes considered in appointment scheduling.

In inpatient care services, patient admission/discharge planning from the hospitalized units is studied [114, 113, 11, 163]. The solutions obtained by SO are structurally different from those solved by other models since SO considers many real-world factors in the model, e.g., patient health status dynamics [11], general LOS distributions [114], and the dependent relationship between the patient LOS and the ICU bed occupancy level [113].

At the online operational level, SO is used to optimize the decisions of ambulance dispatching [26, 176] and ambulance relocation [184, 176]. SO is a suitable method at this level since it is able to mimic the real-time state of the healthcare system. For example, ambulance dispatching is a real-time decision that must take into consideration the current state of the system (i.e. busy and idle ambulances) when selecting the ambulance to respond to incoming emergency calls [26, 176].

In summary, the SO approach is most frequently applied to emergency care services and to strategic decisions (e.g., capacity dimensioning and ambulance location decisions). The SO approach can capture operational dynamics, uncertainty, and complexity in the model. It has the ability to effectively solve large-scale healthcare planning problems. Meanwhile, there is a growing interest to account for behavioral factors of patients and servers in healthcare resource planning, e.g., reneging behavior of patients (i.e., patients deciding to leave without being seen), preferences of patients (e.g., on the service date and on the care staff) and staff working efficiency in relation to different conditions (e.g., the number of waiting patients). Future research could integrate some of these behavioral factors into SO models to have a more realistic representation of the healthcare system.

4 What are the objectives

When the problem characteristics and the key decisions are known to the researchers, they have to address the questions of what the objectives are and how these objectives are handled. Table 4 shows that the patient access times to healthcare services are frequently thought of as an objective in the reviewed research. Here, the access times are a general term which includes many similar measures related to patient waiting, e.g., the typical patient waiting times [64, 86, 182, 168, 119], the number of patients who can be served before/beyond a time limit [170, 1, 35] and the response time by emergency services [177, 26, 37]. Another popular objective is optimizing the LOS in a facility or the patient throughput (i.e., the number of patients that are served). The number of patient deferrals (e.g., patient rejection and patient cancellation) is considered less frequently.

Among the resource-related measures, a main objective is minimizing overtime. Moreover, overtime is often coupled with idle time in the objective function (e.g, [183, 169, 159, 86]), since there exists a clear trade-off between them. A third frequently studied objective is utilization. Leveling the resource availability/workload is sometimes regarded as an objective which basically aims to reduce the variability in the resource use over different days/service units and to achieve a stable operation of the facility. Minimizing the makespan (closing time) of a facility is occasionally considered.

There are some reasons for the SO approach to frequently incorporate the patient access times and/or the resource overtime in the objective function. First, these performance measures might be affected by many environmental factors in the healthcare system and SO is able to effectively model the complex relationship between them. For instance, Woodall et al. [168] and Mousavi et al. [119] resort to simulation (optimization) to estimate the patient waiting time since a closed-form expression is not available in their complex case hospital. Second, the realized values of these performance measures are often not known until the real time, and thus these have to be estimated by simulation. For example, the ambulance response time (i.e., the time between an emergency call and the ambulance

Table 4. Classification based on performance measures in the objective function.

Dimension	Objective	References
Patient-related	Access times	[97, 127, 13, 128, 153, 99, 43, 168, 144, 152, 160, 84, 163, 85, 86, 93, 35, 64, 33, 57, 122, 177, 119, 42, 142, 107, 105, 27, 106, 182, 95, 141, 129, 82, 169, 166, 131, 48, 44, 165, 36, 14, 159, 83, 176, 40, 101, 26, 102, 104, 156, 170, 65, 1, 37, 173, 178, 172, 31, 30, 18, 12, 4, 135, 149, 167, 25, 8, 60, 61, 171, 72, 150]
	LOS	[114, 11, 97, 127, 99, 57, 50, 59, 34, 136, 138, 55, 149, 123, 175, 38]
	Throughput	[97, 99, 128, 43, 105, 3, 115, 88, 149, 167, 117, 157, 164]
	Deferral	[118, 141, 94, 153, 166, 49, 114, 11, 126]
Resource-related	Overtime	[118, 163, 86, 27, 169, 131, 44, 159, 83, 40, 183, 154, 160, 93, 106, 182, 129, 48, 165, 151, 92, 167, 18, 46, 72]
	Idle time	[118, 163, 86, 27, 169, 131, 44, 159, 83, 40, 183, 154, 84, 85, 82, 14, 32, 72]
	Utilization	[97, 99, 128, 127, 50, 94, 153, 49, 13, 113, 47, 149]
	Leveling	[151, 43, 168, 144, 152, 71, 155, 174, 157]
	Makespan	[141, 142, 107, 20, 9, 25, 60]
	Size	[57, 181, 180, 137, 78]
Financial value		[93, 165, 92, 50, 34, 126, 33, 119, 111, 58, 62, 100, 112, 185, 2, 103, 145, 110, 179, 184, 132, 134, 130, 67, 38, 18, 12, 162, 161]
Other		[92, 119, 111, 185, 2, 57, 183, 106, 182, 136, 122, 177, 95, 148, 76, 77, 98, 109, 39, 91, 63, 90, 116, 162, 161, 149, 25, 46, 8, 56, 6, 72, 150, 164]

arrival at the event site [37]) only becomes exactly known when the emergency patient is reached.

Many articles also explicitly consider financial objectives as healthcare facilities need to be cost-efficient when facing limited budgets. Examples include the overall revenue/profit (e.g., [185, 33, 110, 134, 179]), the staffing cost (e.g., [58, 62]), and the financial representation of operational inefficiencies (e.g., overtime and patient waiting time [165, 93]).

Apart from the typical objectives, the SO approach can easily define and measure problem-specific objectives by directly counting them in the simulation output. They are classified into the "other" categorization in Table 4. These include, for example, the fairness and equity measure for care [76, 185], the patient mortality/survival rate [177, 116], and the penalty of service speedup in response to congestion [182].

A vast majority of papers use the form of an expected value in the objective function (i.e., equation (1)). Only a few papers consider other types of measures to optimize, e.g., the variance [43, 152, 155, 153], the maximum [174, 30] and some quantile [48]. Specifically, d'Obrenan et al. [43] minimize the ward occupancy variance during the planning period. Yip et al. [174] minimize the maximum patient occupancy in wards to reduce the peak ward occupancies. Fan et al. [48] develop a robust selection of the best approach and compare it with other approaches based on the mean and the p-quantile of the cost function.

In certain healthcare planning problems, it might be appropriate to consider other types of risk measures. Specifically, quantile-based risk measures are useful since they go beyond the mean and allow healthcare decision makers to determine best systems at various levels of risk. For instance, to reduce patient waiting, conservative decision makers could use a larger quantile for their system selection, while more aggressive decision makers could use a smaller quantile. Also, Coelho and Pinto [37] discuss that it is interesting to study in an emergency medical system not only the average but also the maximum response time. The reason is that minimizing the average response time might reduce the number of ambulances in regions with a small density of emergency calls, possibly increasing the maximum response time in these regions.

SO is a flexible tool in finding high-quality solutions while considering multiple performance measures [83]. Table 5 shows that a majority of studies (around 57%) have multiple objectives to optimize at the same time. Among them, the patient-related (e.g., access times) and the resource-related (e.g., overtime and idle time) performance measures are often simultaneously incorporated in the objective function [44, 159, 86].

When multiple objectives are considered, different types of objective functions have been used to optimize the healthcare planning problem, including three main methods. First, one common method is to use a numerical weight for each objective to convert the multi-objective problem into a single-objective problem (e.g., [160, 182, 159, 166]). This method has some limitations: (1) knowledge is necessary about the weights of the objectives, (2) trade-offs among the objectives are not clear, and (3) it may result in only one final solution. Second, to overcome these disadvantages, an alternative solution consists of searching for the Pareto solutions in the multi-objective problems. Pareto solutions are a set of nondominated solutions, which are chosen, if no single objective could be improved without the sacrifice of at least one other objective. Researchers have developed various optimization algorithms to find Pareto solutions, e.g. non-dominated sorting genetic

Table 5. Classification based on objective function type.

Objective function	Type	References
Single objective		[3, 59, 112, 62, 100, 148, 20, 35, 64, 156, 71, 173, 9, 67, 155, 130, 152, 181, 65, 137, 138, 170, 180, 1, 113, 184, 32, 55, 116, 132, 178, 63, 78, 91, 134, 37, 88, 109, 174, 26, 47, 90, 102, 104, 172, 179, 39, 101, 115, 56, 6, 123, 175, 167, 117, 31, 30, 4, 61]
Multiple objectives	Weighted	[93, 92, 84, 85, 97, 86, 163, 43, 58, 160, 14, 98, 151, 82, 83, 42, 145, 77, 99, 141, 154, 27, 131, 105, 183, 110, 106, 118, 165, 36, 40, 44, 182, 48, 136, 159, 166, 176, 129, 169, 162, 161, 149, 25, 46, 8, 18, 12, 60, 135, 72]
	Pareto-based	[111, 114, 107, 142, 34, 153, 50, 57, 185, 49, 2, 76, 103, 119, 95, 126, 11, 167]
	Stage-based	[160, 168, 94]

algorithm II (NSGA-II) [2, 126, 50, 50] and ϵ constraint method (to constrain one objective function) [114, 11, 185]. Third, several studies address the problem with multiple objectives in different stages in which a series of single-objective problems are optimized separately [94, 168, 160].

In summary, SO has the ability to include various (typical and problem-specific) performance measures in the objective function. Due to this ability, future research could integrate the effect of behavioral factors into the objectives in SO models, e.g., the penalty of a service speedup. However, the challenge is how to accurately measure the relationship between the behavioral factors and the objective outcomes, which might be nonlinear and complicated in healthcare. In the objective function, it is interesting to consider variance- or quantile-based risk measures in healthcare. Some advanced SO literature has provided evidence supporting the use of quantile estimates as the measure of comparison (e.g., [19]). In addition, identifying Pareto solutions in multiple-objective SO (MOSO) problems in healthcare is an interesting area since preferences among healthcare stakeholders are always conflicting [133]. The challenge is that solving a MOSO problem normally requires more computational effort than solving single-objective problems. According to a recent review [74], the future development of MOSO methods should notice two points: (1) keeping parallel algorithmic implementation in mind, and (2) first studying biobjective problems.

5 What are the SO approaches

In order to achieve the desired goal, researchers should now decide on how to model and solve such problems in terms of simulation methods and of optimization algorithms/packages.

5.1 *Categorization based on simulation method*

Table 6 categorizes the articles based on the simulation modeling methods. In the reviewed papers, discrete-event simulation (DES) is by far the most popular method to model healthcare operations. The reason is that DES enables the study of discrete, dynamic, and stochastic systems with various levels of detail and complexity in the healthcare environment. Monte Carlo simulation (MCS) is the second most popular method. MCS is a discrete, static, and stochastic simulation method, which is a scheme applying random numbers for solving certain problems. MCS is usually used to estimate the unknown function (i.e., $f(x, y, \omega)$ or $g(x, y, \omega)$) by repeatedly generating many independent sample paths for each solution. Agent-based simulation (ABS) is found to be the least popular simulation method from the three. However, ABS can simulate the interacting and autonomous agents whose behaviors and interactions with their environment (including other agents) are modeled as a set of behavioral rules [77]. For this reason, we observe that ABS is a useful tool to study healthcare systems with complex behaviors, e.g., the planning of epidemic-control resources [98, 77] and the planning of hospital emergency departments [97, 109].

Table 6. Simulation modeling methods.

Method	References
DES	[185, 91, 84, 83, 37, 145, 26, 78, 180, 115, 152, 14, 181, 122, 119, 76, 64, 62, 104, 106, 63, 50, 2, 126, 34, 103, 153, 49, 107, 142, 111, 128, 101, 102, 134, 35, 166, 129, 183, 178, 110, 39, 90, 136, 184, 131, 173, 59, 141, 13, 179, 27, 82, 9, 1, 169, 132, 57, 114, 168, 144, 127, 42, 36, 95, 137, 113, 138, 47, 174, 11, 55, 65, 33, 67, 38, 149, 60, 135, 56, 123, 117, 31, 30, 4, 61, 157, 171, 164]
MCS	[3, 40, 176, 94, 92, 118, 93, 154, 43, 109, 25, 18, 150, 129]
ABS	[109, 98, 97, 99, 77, 6, 175]

Notes. DES: Discrete-event simulation; MCS: Monte Carlo simulation; ABS: Agent-based simulation.

5.2 *Categorization based on paper focus (algorithm/package)*

Table 7 provides a classification of the literature based on three categories: methodology-focused, domain-focused with developed algorithms or with optimization packages. Methodology-focused papers focus on developing/improving SO algorithms and usually use simple healthcare resource planning problems (so-called toy problems) to demonstrate the performance of the proposed SO algorithm. In contrast, domain-focused papers aim to tackle practical healthcare resource planning problems (with the help of SO), and they tend to relax some stringent assumptions or to involve more realistic characteristics in the SO model. To solve such problems, some authors might develop and apply efficient algorithms (i.e., domain-focused with developed algorithms), while others might directly apply existing optimization packages (i.e., domain-focused with optimization packages).

Table 7 reveals that the focus of SO is on solving practical healthcare problems while methodology-focused papers with healthcare "toy problems" are very rare. Although a majority of articles develop/apply sophisticated SO algorithms to solve their planning problems, a fair number of domain-focused SO studies also use built-in optimization packages within a simulation software. OptQuest is the most popular optimization package which is often integrated with Arena, a DES software (e.g., [179, 27, 11]). In addition, some algorithm-focused papers (e.g., [37, 154, 178]) use OptQuest as a comparison/benchmark.

Table 7. Classification based on focus.

Type	Focus	References
Developed algorithm	Methodology	[112, 48, 107, 170]
	Domain	[154, 122, 110, 98, 184, 101, 131, 129, 145, 160, 93, 3, 26, 116, 76, 90, 43, 148, 136, 173, 142, 39, 64, 20, 118, 91, 104, 159, 78, 119, 151, 50, 155, 180, 153, 62, 111, 182, 183, 115, 152, 77, 14, 99, 97, 126, 130, 84, 92, 141, 35, 176, 63, 166, 134, 2, 100, 49, 106, 32, 105, 34, 59, 37, 185, 103, 102, 44, 181, 177, 94, 163, 83, 165, 58, 40, 178, 156, 172, 88, 162, 161, 8, 12, 6, 175, 25, 18, 60, 135, 56, 123, 31, 4, 46, 61, 72, 150, 164]
Optimization package	Domain	[154, 122, 110, 113, 109, 85, 114, 138, 86, 71, 179, 13, 47, 168, 137, 174, 144, 127, 95, 132, 42, 27, 11, 82, 169, 57, 36, 9, 55, 1, 65, 33, 128, 37, 178, 46, 38, 149, 117, 30, 157, 171]

5.3 *A gentle introduction to simulation optimization algorithms*

We classify the SO algorithms according to whether they can address problems with discrete or continuous decision variables [17, 5], as shown in Table 8. The acronyms associated with the SO algorithms will be used throughout this paper.

Table 8. Classification of SO algorithms and their associated acronyms.

Variables	Solution space	Algorithms	More specific algorithms
Discrete decision variables	Finite and small solution space	Ranking and selection (**R&S**)	Indifference-zone procedures (**IZ**) Optimal computing budget allocation (**OCBA**)
		Multiple comparison procedures (**MCPs**)	
	Large/infinite solution space	Ordinal optimization (**OO**)	
		Direct search methods (**DSMs**)	
		Random search methods (**RS**)	Tabu search (**TS**)
Continuous decision variables			Genetic algorithms (**GAs**) Simulated annealing (**SA**) Scatter search (**SS**) Particle swarm optimization (**PSO**)
		Metamodel-based methods (**MMs**)	Response surface methodology (**RSM**) Artificial neural networks (**ANNs**) Kriging (**KR**)
		Gradient-based methods (**GMs**)	
		Sample path optimization (**SPO**)	

Notes. Many of the algorithms that are applicable to the problem with continuous decision variables are also, with some modifications, applicable to the problem with large/infinite solution spaces, e.g., ordinal optimization, direct search methods, and random search methods.

When decision variables are discrete and the solution space is finite and small (typically with hundreds of solutions), algorithms include ranking and selection (R&S) and multiple comparison procedures (MCPs). R&S has two main formulations: (1) to minimize the number of simulation runs needed subject to a specified probability of correct selection of solutions (most procedures try to guarantee that the performance of the eventually selected solution differs from the optimal solution performance by at most δ, called the indifference-zone (IZ) [79]) and (2) to maximize the probability of a correct selection subject to a specified computational budget (called the optimal computing budget allocation (OCBA) [29]). Unlike R&S, MCPs run a number of simulation replications per solution to draw inferences on the performance measures by constructing confidence intervals.

When the feasible solution space is large (or potentially infinite), algorithms that have a search procedure are required. Ordinal optimization focuses on sampling a selected subset of the solutions and on evaluating them to choose the best one [68]. Direct search methods are defined as the sequential examination of potential solutions generated by some strategy [89]. Many direct search methods developed for SO are extensions of the ideas for derivative-free optimization [89]. Random search methods sample a set of feasible solutions in each iteration, conduct simulations, and then use the simulation results to decide what solutions should be sampled in the next iteration and which one is the current best solution [7].

In the area of continuous SO problems, metamodel-based methods are typically useful and focus on learning the relationship between input decision variables and output responses to approximate a potential functional form (known as a metamodel) [17]. The functional forms can be diverse, e.g., polynomials, artificial neural networks, and Kriging [15]. Response surface methodology is a metamodel-based method that often constructs linear/quadratic approximations to the objective function [17]. Unlike metamodel-based methods, gradient-based methods and sample path optimization require both differentiability and continuity in the objective function. Gradient-based approaches attempt to make a movement in the gradient direction given a current best value of the decision variables [52]. Sample path optimization first takes many simulation runs, and then tries to optimize the resulting estimates, rather than working with the underlying unknown function itself.

5.4 Categorization based on SO algorithm

5.4.1 *Categorization based on variables and stochastic constraints*

According to Table 9, most papers focus on the problems that only have discrete decisions to make. These discrete decision variables might either be binary (e.g., patient-to-appointment scheduling [40, 44]), integer-ordered (e.g., the number of staff [58, 35, 88]), or categorical variables (e.g., indexing different combinations of multiple input parameters [128, 57, 159]). A few papers solely involve continuous decision variables, e.g., time-related decisions (e.g., the appointment time of patients [182, 154, 20]) and budget-related decisions (e.g., [55, 77]). Relatively fewer papers address both discrete and continuous variables together. These papers tackle the two kinds of variables either in a separate way [160, 136, 126, 110], by assuming some discrete variables to be continuous and rounding the optimal solution to get integer values [166, 111], or by solving a mixed integer linear programming [185, 118].

In some SO problems, certain constraint functions might not be directly available, referred to as stochastic constraints (see Table 10). Specifically, they are constraints imposed on the maximum expected waiting times (e.g., [185, 100, 85]), on the probability of patient rejection [114, 11], on the expected overtime [43, 93], and on the expected bed occupancy [126, 43]. Three methods are observed in the reviewed papers to address the stochastically constrained problem. First, stochastic constraints can get into objective functions by using penalty functions [126, 111, 35] or a Lagrangian function [112], which facilitates the use

Table 9. Classification based on the type of decision variables.

Type of variable	References
Discrete	[128, 169, 165, 172, 109, 100, 114, 113, 144, 11, 13, 177, 33, 43, 176, 40, 163, 92, 93, 97, 99, 129, 48, 58, 44, 86, 106, 131, 27, 85, 141, 84, 42, 83, 82, 98, 14, 94, 168, 32, 1, 101, 119, 39, 115, 103, 64, 35, 26, 104, 2, 76, 90, 102, 179, 37, 174, 95, 148, 57, 88, 62, 50, 34, 78, 116, 142, 178, 132, 184, 138, 180, 137, 107, 170, 112, 181, 152, 67, 3, 173, 71, 156, 151, 159, 145, 46, 38, 149, 117, 30, 167, 162, 161, 8, 12, 6, 175, 135, 123, 31, 61, 72, 164]
Continuous	[122, 36, 182, 183, 105, 77, 154, 20, 55, 9, 47, 134, 153]
Discrete & continuous	[160, 118, 166, 136, 110, 126, 185, 111, 25, 18, 56]

of a standard SO algorithm to solve the problem. The major challenge of using penalty functions is the proper tuning of penalty parameters. Second, some SO algorithms (e.g., R&S [160, 100] and sample path optimization [163, 92, 43, 93]) themselves or with suitable adaptions have the ability to deal with the stochastic constraints. Third, specific techniques (e.g., violation-constraint handling [64]) can be incorporated into the SO algorithm.

Table 10. Classification based on the stochastic constraints.

Object	Type	References
Patients	Expected waiting time /service quality	[58, 3, 59, 112, 85, 97, 62, 100, 20, 35, 64, 160, 185, 25, 18, 30, 31, 167]
	Expected rejection /cancellation rate	[114, 11, 163]
Resources	Expected overtime	[93, 92, 163, 43, 160, 167]
	Expected occupancy	[43, 163, 126]
Other		[84, 111, 110, 18]

Notes. Other constraints include, e.g., caregivers' duty length.

5.4.2 *Categorization based on used SO algorithm*

Table 11 categorizes the papers according to the SO algorithms. It is observed that random search methods have been by far the most popular SO technique applied. Genetic algorithms are the most frequently used random search methods. Generally, the GA-based SO includes four steps: (1) initialize a population of solutions (chromosomes), (2) evaluate the solutions (fitness) through simulation, (3) choose solutions and generate a new solution (offspring) by crossover and mutation, and (4) repeat until certain stopping criteria are met. To further improve the problem solving efficiency of GAs, additional approaches such as R&S (OCBA [63, 153, 49] and IZ [165]) and other RS methods (e.g., simulated annealing [63] and custom neighborhood search algorithm [185]) have been integrated with GAs. Tabu search shows up as another popular random search method in the SO framework. Similarly to GAs, simulation is used for evaluating the move of the TS iteration to reach better solutions. To improve TS's efficiency, some researchers adapt the basic TS [177, 2, 142]. For example, a reactive tabu search is developed to solve the ambulance location problem modeled in the SO framework [177].

Table 11. Classification based on applied SO algorithms.

Algorithm	Type	References
R&S	IZ	[160, 100, 165, 48]
	OCBA	[34, 101, 102, 64, 50, 63, 35, 153, 49, 103]
MCPs		[170]
OO		[103, 134]
DSMs		[110, 39, 111]
RS	GAs	[90, 184, 131, 116, 148, 173, 153, 49, 59, 107, 63, 50, 2, 126, 185, 175, 25, 18, 4, 56, 46]
	TS	[92, 94, 43, 136, 142, 84, 141, 32, 105, 177, 83, 40, 2, 123]
	SA	[92, 136, 91, 35, 63, 167, 60]
	PSO	[34, 101, 102, 12, 31]
	SS	[90, 84, 32, 83]
	Other	[92, 94, 177, 40, 118, 163, 98, 76, 64, 104, 119, 99, 106, 172, 166, 159, 62, 6]
MMs	RSM	[178, 122, 105, 154, 77, 88]
	ANNs	[64, 119, 178, 38, 175, 135]
	Kriging	[37, 12]
	Other	[178, 130, 166, 156, 185, 6]
GMs		[166, 122, 160, 129, 182, 183, 112]
SPO		[160, 92, 94, 118, 163, 93, 176, 43, 20, 8]
Other		[84, 83, 176, 37, 156, 40, 98, 185, 91, 145, 3, 26, 20, 78, 151, 155, 180, 115, 152, 14, 97, 44, 181, 58, 27, 12, 31, 4, 135, 162, 161, 150, 61]

Notes. Other algorithms include, e.g., custom heuristics for optimization (e.g., [115, 176, 58, 180]) and additional techniques to improve the classical SO algorithms (e.g., neural network accelerators [84, 83] and variance reduction techniques [40, 27]).

Metamodel-based methods, among which RSM and ANNs are the most popular methods, are next in line when enumerating frequently used SO algorithms. Most of the RSM-based articles use two sequential procedures to approximate a polynomial metamodel (e.g., [122, 154, 77, 105]). ANNs are powerful to estimate complex nonlinear input-output relationships [178, 119, 64]. The primary drawback might be the excessive use of simulation runs in one area before exploring other areas of the search space [53]. Kriging is a more efficient method since it provides at candidate solutions a predictor variance which can indicate poorly sampled areas in the solution space. For example, Coelho and Pinto [37] apply Kriging-based SO to solve the ambulance location problem and the predictor variance is used to balance local and global searches.

Sample path optimization is often integrated with other optimization methods (e.g., RS methods) to find (near-)optimal solutions in the reviewed papers (e.g., [93, 118, 94, 92]). Compared to SPO, gradient-based methods using sample gradients are generally more efficient [183, 160]. Although the gradient-based methods are typically designed for SO problems with continuous input variables, there exists research that uses it to deal with discrete input variable problems [166, 129, 112]. For instance, in Pan et al. [129], discontinuities of the gradient cost function are smoothed by smoothed perturbation analysis to derive the unbiased gradient estimation.

Among the R&S methods, basic IZ procedures are usually used to select better solutions (e.g., appointment/surgery schedules [165, 48, 160], staffing levels [100]) among a small set of solutions with some statistical guarantee. Extensions of IZ frameworks are introduced in some research [48, 165]. Specifically, Fan et al. [48] develop a new IZ formulation and design two selection procedures that can select the solution with the best worst-case mean performance in the presence of input uncertainty (i.e., robust simulation optimization). Another R&S method, OCBA, is usually integrated with random search methods in order to effectively allocate simulation budgets in the search process for single-objective problems [101, 64, 102] or for multi-objective problems (in this case, MOCBA) [34, 49, 50].

Many studies advise to consider a more nuanced aspect of healthcare operations by incorporating important real-life features that have been ignored. As a result, existing algorithms might become more time-consuming for large-scale and complex problems. Future research needs to further improve the efficiency of algorithms. One possible approach is to devote more computational power to identifying elite solutions within early iterations and then to increase the accuracy of the best solution for later iterations. Another challenge consists of specifying an appropriate input distribution in simulation modeling since there might be multiple possible distributions to fit the input parameters. Therefore, robust simulation optimization methods could be developed to deal with this situation, which seek solutions that are robust to the input uncertainty. Two main approaches exist for robust SO: the worst-case approach (see [21]) and the mean-variance trade-off approach (see [41]). Robust SO has seldom been investigated in the healthcare resource planning context (we are only aware of [48, 12]), which constitutes an interesting area for future research.

6 Validating and applying the model

The final steps consist of validating the developed SO models (in terms of the simulation component and the whole model) to illustrate the applicability of the research. A cornerstone to the success of applying SO to healthcare planning is to check whether the simulation model can provide a satisfactory range of accuracy that is consistent with the intended application of the model (i.e., model validation [143]). As evident from Table 12, only 41% of the surveyed articles explicitly mention performing a validation of the simulation model, while many articles do not mention this. Several ways are used for model validation in the literature.

First, if the real data are available, model validation is often performed by comparing the simulation output with the existing system data regarding a single performance measure (e.g., [61, 13, 36, 88]) or multiple performance measures (e.g., [47, 35, 64, 179, 90, 97]). Normally, simultaneously comparing multiple performance measures can provide a more accurate simulation model, since many healthcare planning problems consider multiple components of the real system and involve several performance measures. The comparison could be done informally by simply comparing summary statistics (e.g., the sample mean [35, 90, 13, 100]) or more formally conducted by performing appropriate statistical procedures (e.g., a correlation test [90, 34], a hypothesis test [36, 149], and confidence interval procedures [173, 178, 172, 88, 128]) for the performance measures.

A formal statistical procedure is more reliable than simply comparing summary statistics since the simulation outputs are stochastic. Whereas a hypothesis test produces only a "reject"/"fail-to-reject" conclusion, a confidence interval procedure provides this information (according to the interval missing/containing zero) and further quantifies the magnitude by which the expectations differ. The latter procedure is more preferable for model validation since the users can decide whether the difference between the model and the

Table 12. Validation of simulation models and test of SO models.

	Type	References
Validation of simulation models	Real data (single measure)	[157, 36, 128, 91, 55, 116, 61, 172, 13, 37, 88, 31, 33, 32, 113, 141, 138, 107, 65, 30, 14, 150]
	Real data (multi-measure)	[95, 57, 34, 63, 1, 168, 67, 173, 117, 47, 35, 64, 179, 90, 49, 135, 175, 100, 149, 153, 111, 178, 97, 137, 123, 162, 161, 157]
	Expert opinion	[36, 95, 57, 128, 34, 63, 91, 55, 116, 1, 168, 67, 173, 2, 159, 174, 50, 157]
	Others (unclear)	[149, 4, 144, 78, 110, 77, 98]
Test of SO models	Theoretical data	[9, 83, 119, 183, 118, 112, 176, 154, 40, 20, 92, 148, 93, 78, 18]
	Real data	[42, 180, 103, 26, 115, 145, 184, 76, 129, 39, 132, 27, 82, 169, 114, 11, 122, 134, 152, 181, 142, 62, 102, 136, 84, 185, 3, 101, 126, 104, 106, 166, 59, 131, 156, 94, 163, 99, 170, 48, 109, 71, 85, 86, 151, 155, 44, 58, 177, 160, 182, 130, 165, 43, 110, 95, 57, 127, 65, 33, 179, 1, 36, 137, 168, 144, 113, 138, 174, 55, 13, 47, 37, 91, 35, 64, 90, 141, 111, 14, 50, 34, 128, 107, 173, 67, 178, 153, 49, 2, 172, 77, 88, 98, 97, 32, 105, 116, 159, 100, 46, 4, 25, 56, 167, 8, 38, 123, 12, 31, 162, 161, 6, 175, 30, 149, 117, 135, 60, 61, 157, 171, 72, 150, 164]
Implement-tion	Support tool	[127, 174, 20, 128, 178, 168, 181, 3, 98, 38, 161, 164]
	In practice	[101, 103, 97, 138, 99, 164]

system is practically significant, depending on the model's purpose and the user's utility function. Furthermore, the assessment and comparison could be made by using graphical plots, which are quite useful for facilitating understanding and buy-in from practitioners.

Second, expert opinion and experience could be taken into account in the development of the simulation model. This facilitates the simulation model to be reasonable and consistent with perceived system behavior (called face validity). In the reviewed literature, this way is usually used as an additional and reinforcing step to validate the model (e.g., [63, 168, 128, 116, 36]). There is often "a tacit assumption that statistical methods for comparing model output with observed data are the "gold standard" and face validity is an inferior method of validation" [23]. However, when the real data are not available, it is important to incorporate expert knowledge and to discuss simulation results with healthcare practitioners (e.g., [159, 174, 50, 157]).

The whole SO model should be tested with data and/or applied in practice to examine whether the model performs as expected. Table 12 shows that a majority of papers test their research based on real-world data. However, simply testing developed models on real-world data does not mean that the models can be implemented in reality. We define a two-level scale of implementation: (1) the development of a decision support system/graphical user interface, and (2) the actual implementation in practice. The former aims to develop a user-friendly decision support tool based on the proposed SO model for healthcare managers to use in practice, while the latter describes the actual test or implementation of the proposed model in a real healthcare facility. Notice that even properly built decision support systems may not be adopted/implemented in practice. We learn that less than 12% of the articles (around 8% for (1) and 4% for (2)) report on the implementation of the models. The latter figure is comparable to previous findings in the healthcare OR/MS literature review of Brailsford et al. [24] who report a figure of 5.3%. The reasons for the small number of implementations in practice might be the difficult development of a platform system that needs to be linked to many healthcare databases [127] and the lack of enough data that can be collected [63, 91, 109]. Moreover, the developed methods might be too complicated to implement in practice and the methods require a large change in provider practice/culture.

Among those sparse studies that report on actual implementation, Li et al. [101] conduct a pilot study in the hospital endocrine department to validate their SO-based research that manages the service priorities between appointment patients and walk-in patients in real time. Woodall et al. [168] indicate that some of their solution results for staffing decisions at a cancer center have been adopted by the management. However, these authors provide few details about the implementation process and the lessons learned from it. In contrast, Lee et al. [97] provide detailed steps to implement the developed SO system in a hospital emergency department for optimizing the workflow. They learn that the collaborative effort between healthcare participators and OR researchers results in scientific advances on both fronts. Similarly, Visintin et al. [164] present some fundamental aspects to the successful implementation, mainly including (1) to reduce top management and staff skepticism about the model and (2) to introduce changes to their established way of working.

In short, SO models appear to suffer from the same issues as other healthcare OR/MS approaches, i.e., only a small part of the papers reports on the (real-life data based) model validation of simulation and even fewer papers report on the implementation in practice.

For future work, we encourage the development of accurate healthcare simulation models and the provision of key information on the actual implementation. To facilitate the implementation, it is important that the proposed method should be easy-to-implement (and -understand) on the practitioner side at least to some degree. The practical implementation may also require the development of Information Communications Technology solutions that are possibly integrated with the healthcare organizations' database systems.

7 Relations between classification fields

In this section, we will look at the relations between the individual classification fields. Tables 13 and 14 combine a few fields and list numbers of publications at the intersections.

Table 13 (from the column perspective) shows that the DES method and RS algorithms have extensively been used with almost all the fields. MCS and SPO are frequently applied to similar fields. Both are more likely to be applied to lower-level decision problems (i.e., operational offline level), where stochastic constraints and multiple objectives are involved. This is understandable as MCS is often used to estimate mathematical expectations of objective functions (e.g., resource overtime) and of stochastic constraints.

ABS is often applied to strategic decision problems with discrete decision variables. This seems counterintuitive since ABS is able to handle lower-level decision problems. One reason is that ABS is exclusively used to study emergency care services in the reviewed literature [97, 109, 98, 77]. As Hulshof et al. [73] indicated, emergency care services inherently have shorter planning horizons than other care services. As a result, some of the strategic decision problems of emergency resource planning are much related with lower-level operations that could be modeled in ABS. Furthermore, in ABS-based papers, the developed SO algorithm is exclusively tested on real-life data and the results have a higher likelihood (3 out of 7 papers) to be implemented in practice compared to other simulation methods. Thus, it is emerging and promising to integrate ABS into different SO algorithms to facilitate the implementation. However, ABS is seldom integrated with optimization software packages.

Table 13 also shows that MMs are another popular SO algorithm that can be applied to various problem settings and combined with different simulation methods. The relatively lower numbers of publications compared to RS (in many fields) indicate that there are many opportunities for future work. R&S, MCPs and OO are often developed to deal with

Table 13. The number of publications to perform a method/implementation (column) with a field (row).

Field		Simulation method			SO algorithm						Implement.
	All	DES	MCS	ABS	R,M, O*	DSMs	RS	MMs	GMs	SPO	
All		*89*	*14*	*7*	*16*	*3*	*71*	*18*	*7*	*11*	*17*
Strategic level	*64*	44	2	6	8	2	27	13	2	0	11
Tactical level	*32*	25	0	0	3	0	20	2	1	0	5
Offline level	*38*	15	10	0	4	0	21	2	4	10	1
Online level	*9*	7	2	0	0	0	4	1	1	1	0
Discrete var.	*104*	64	9	6	11	1	53	11	2	8	15
Continu. var.	*13*	8	1	1	2	0	4	4	3	1	1
Both	*11*	7	3	0	1	2	8	2	2	2	0
Stoch. constr.	*29*	13	6	1	4	2	17	2	2	6	3
Single obj.	*59*	41	2	3	8	1	28	8	1	1	7
Multiple obj.	*71*	38	11	4	8	2	41	8	5	10	6
DES	*89*	89	1	0	11	3	47	9	4	0	11
MCS	*14*	1	14	1	0	0	7	1	1	7	1
ABS	*7*	0	1	7	0	0	4	3	0	0	3
Method. focus	*4*	1	0	0	2	0	1	0	1	0	0
Doma. develop.	*103*	56	13	6	16	3	56	17	7	11	11
Doma. package	*42*	36	2	1	0	1	15	5	1	0	7
Validation	*59*	46	1	4	8	2	29	7	0	0	8
No mention	*82*	43	13	3	8	1	42	11	7	11	9
Theory data	*15*	5	7	0	0	0	8	2	2	6	1
Real data	*125*	83	7	7	15	3	62	16	5	5	16
Implement.	*17*	11	1	3	2	0	4	2	0	1	17

Notes. *All* means the total number of publications associated with that row or column. *All* might not equal the sum of the numbers in that row/column since multiple methods might be used in a paper or "other" methods are used. *R, M, O: R&S, MCPs, OO.

decision problems with discrete variables and integrated with DES in a real-data context. We find a strange trend that all GMs-based papers do not explicitly mention the validation procedure of simulation even though those papers are mainly domain-focused. One reason might be that those papers focus more on developing and testing advanced SO algorithms.

In addition, implementation is more likely to be reported in papers that handle higher decision levels (i.e., strategic and tactical levels) with discrete variables and that use real-life data to test the models. This might be explained by the fact that higher decision levels involve a smaller number of stakeholders of the healthcare facility (e.g., top management) and avoid too many interruptions to the day-to-day working of frontline staff.

Table 14. The number of publications to use an objective (column) with a field (row).

	All	Objective: Single	Objective: Multi.	Performance measures: Acce. times	LOS	Throu-ghput	Def-erral	Over-time	Idle time	Utili-zation	Leve-ling	Make-span	Finan-cial
All		*59*	*71*	*73*	*16*	*13*	*6*	*25*	*18*	*12*	*9*	*7*	*29*
Strategic level	*64*	30	25	26	12	8	3	0	0	8	2	1	22
Tactical level	*32*	13	17	23	1	3	0	9	11	0	5	0	2
Offline level	*38*	9	29	22	3	2	3	17	8	4	2	6	4
Online level	*9*	7	2	5	0	1	0	1	0	0	0	0	2
Stoch. constr.	*29*	12	18	15	4	4	3	7	4	1	1	2	10
DES	*89*	41	38	46	13	6	6	7	8	9	5	5	18
MCS	*14*	2	11	8	0	2	0	7	3	1	1	1	3
ABS	*7*	3	4	2	3	2	0	0	0	2	0	0	0
DSMs	*3*	1	2	0	0	0	0	0	0	0	0	0	2
RS	*71*	28	41	41	7	5	4	14	13	7	1	6	13
MMs	*18*	8	8	10	2	2	1	1	1	0	0	0	5
GMs	*7*	1	5	5	0	0	1	4	1	0	0	0	1
SPO	*11*	1	10	7	0	1	0	6	3	1	1	1	2
Validation	*59*	31	22	29	9	6	2	1	3	9	4	2	11
No mention	*82*	28	49	44	7	7	4	24	15	3	5	5	18

Notes. *All* means the total number of publications associated with that row or column.

From the row perspective of Table 13, we notice some interesting connections. First, many of the SO algorithms are applicable to the other type of decision variables and to the stochastically constrained problem, with some modifications. This means that the type of decision variables (discrete or continuous) and the incorporation of stochastic constraints are not decisive criteria when choosing the SO algorithm. Second, multi-objective problems tend to be more common than single-objective problems under almost all SO algorithms. This reveals the capability of the SO approach to deal with multiple objectives.

Table 14 reveals some interesting patterns between the objective function field and other classification fields. First, the decision problems of tactical and offline operational levels are more likely to consider multiple objectives (e.g., patient access times, resource overtime and idle time), while this is not the case for strategic decision problems where patient access times or financial objectives are often considered. Second, stochastically constrained papers often involve the objectives of patient access times, financial values and resource overtime. Most of these measures also appear in the stochastic constraints (see Table 10). For example, patient access times can be optimized in the objective function or can be defined in a constraint to limit the maximum times. Third, validation-mentioning literature often uses the objectives of patient access times, LOS, utilization, throughput and financial measures. Understandably, these measures (especially, the first four) are often used as key performance indicators to validate the simulation model in the literature.

8 Discussion

8.1 Where to use SO for healthcare

The number of articles on using SO to study healthcare resource planning problems in recent years is obviously growing (from 19 articles in 2004–2012 to 117 articles in 2013–2021, see Figure 1). Despite this fact, the relative use of SO to other OR/MS methodologies (e.g., simulation, mathematical programming and analytical stochastic models) in similar healthcare problems still remains small. Specifically, simulation modeling has been used for a wide range of applications in healthcare [22, 24]. Mathematical programming (e.g., linear programming and integer programming) and analytical stochastic models (e.g., queueing theory and Markov Decision Processes) are also popular methods to deal with resource planning in the six care services according to Hulshof et al. [73].

There are several advantages of using SO and suitable situations to deal with complex healthcare resource planning problems compared to other methods. First, SO techniques inherit specific advantages of the traditional simulation method in modeling healthcare systems [22]. Meanwhile, SO can overcome the limitation that occurs in a typical simulation study, where only a limited number of system alternatives are evaluated and compared. This is one of the frequently recognized benefits in the literature (e.g., [83, 64, 86, 107, 57]) since healthcare resource planning often involves a large solution space (see Section 3).

Second, SO is suitable for solving an optimization problem when the objective function has no analytical closed-form expression, which is common in complex and stochastic healthcare systems [168, 119, 134, 185, 107]. Mathematical programming techniques have a limited ability for this kind of problems. Although stochastic mathematical models (e.g., stochastic programming) could overcome some of these limitations, they almost invariably make the model still more complicated [115, 160]. For instance, Marla et al. [115] discuss that it is more difficult to incorporate the abandonment behavior of patients in stochastic programming than in an SO framework.

Third, SO is more tractable and suitable for large problem instances than mathematical programming models [182, 160, 78]. For example, Zhang et al. [182] formulate the appointment scheduling problem as two formulations: a stochastic integer programming model and a simulation optimization model. They show that the former model can exactly solve moderate-size problem instances under certain assumptions and the latter achieves near-optimal solutions for the same problems, while the latter is also able to scale up to much larger problem instances.

Fourth, analytical stochastic models such as queueing theory and Markov Decision Processes could also solve healthcare planning problems under uncertainty [73]. The main difference between simulation (optimization) and analytical stochastic models is that the former can describe the detailed activities of individuals and the interactions between them, while the latter primarily study the flow of homogeneous groups with more strict problem assumptions (e.g., certain probability distributions). If it is necessary to model a healthcare system at a higher level of complexity and detail, SO is more suitable [83, 44, 11, 90, 91, 116, 165]. In addition, an SO model is suitable for the situation in which the objective function cannot be effectively approximated by analytical models [166, 174, 152]. For instance, if the objective is to reduce the variability in the scheduled workload, SO may be suitable since the daily workload variability is not easy to capture in analytical models.

However, SO has several disadvantages, including the time, effort and data involved in developing, validating, and then experimenting the model. Nevertheless, the time and effort required for model building can be certainly decreasing as more researchers gain experience and as the power of available software packages increases. The cost of collecting the required data would ease with the ever-improving healthcare information systems. In general, a useful guideline in SO modeling is to keep the model as simple as possible while capturing all necessary factors (key inputs and objective functions) of interest.

8.2 How other methods complement SO for healthcare

An effective resource planning in healthcare might require a combination of multiple methodologies, which can take advantage of their complementary nature. In this section, we expand the discussion to other approaches to healthcare planning and their relation to SO to provide useful insights on how different methodologies could complement SO.

Analytical stochastic models may be used in an exploratory way to determine the effects of demand and supply on resource use. The advantage is that once programmed, the models can be tested and run quickly. In particular, the analytical stochastic models complement SO for healthcare planning in three ways.

First, they can provide valid/theoretical insights on good modeling assumptions, e.g., how patient arrivals [39] or service times [63, 91] should be modeled at the emergency department, and how the patient flow should be modeled [90, 176, 11, 181]. The fact that simulation allows for more realistic modeling assumptions does not resolve all modeling difficulties, e.g., the tradeoff between model confidence and tractability. The analytical stochastic models in healthcare planning provide insights for managing this tradeoff, which would be useful for developing high-confidence simulation models and tractable optimization approaches. Also, many impactful papers outside the reviewed literature use analytical models to provide theoretical insights on good modeling choices, e.g., on patient arrivals [80, 81], on service times [146, 45] and on patient flows [70] at a healthcare facility.

Second, they can identify structural insights for healthcare planning problems, which can then make the problem amendable to SO (e.g., [102, 114, 104, 176, 181, 40]). For instance, if a paper identifies the structure of the (near-)optimal policy of a dynamic healthcare planning problem to be of threshold-type (e.g., [147]) or index-based (e.g., [28]) using a simplified model, this insight can then be used in combination with a simulation model that relaxes the assumptions and uses SO to find the optimal threshold or to apply the index policy for the more realistic model. In addition, Li et al. [102] discuss that the investigation of a simplified patient referral process (with a threshold policy) between hospitals based on an analytical queueing model can help to identify promising directions of deriving the optimal threshold for the general model in the SO method.

Third, they can provide approximate formulas on the performance measures, which can increase the computational efficiency in SO (e.g., [103, 134, 119, 160, 166]). For example, Mousavi et al. [119] calculate the time of transferring the injured patients to trauma centers by queueing theory (due to available analytical formulas), and calculate the waiting time spent inside the trauma center by simulation. Li et al. [103] and Qiu et al. [134] apply a multi-confidence model-based optimization approach, which integrates the advantages of queueing model (low confidence, but quick estimation) and simulation (high confidence, but time-consuming). This integration can relieve the computational challenges in SO for healthcare.

Mathematical programming techniques are usually used to provide (near-)optimal solutions subject to a set of constraints that portray the conditions under which the decisions have to be made. The optimal solutions are searched in an intelligent manner, allowing the evaluation of healthcare planning problems with a large feasible solution region. Mathematical programming models could complement SO for healthcare planning in two ways.

First, they can be embedded into the SO framework for healthcare planning to solve a subproblem or to help to find better solutions [168, 141, 43, 142, 151, 176, 83]. For example, Klassen and Yoogalingam [83] embed mixed-integer programming in the optimization step of SO to help quickly approach an optimal solution. Saremi et al. [141] incorporate integer programming and binary programming in SO, which helps to identify the initial solutions and to provide promising solutions. In d'Obrenan et al. [43], mathematical programming is first used to solve the healthcare planning problem and then the solutions are further improved by SO. Furthermore, methods developed for mathematical programming models can be applied to find optimal solutions of SO once the metamodel is obtained [185, 88].

Second, they can identify structural insights for healthcare planning problems, which can provide a foundation for applying SO [83, 86, 141, 160, 182]. As noted by Klassen and Yoogalingam [83], some stochastic mathematical programming studies for appointment scheduling reveal that the optimal appointment intervals exhibit the dome pattern. This insight may be used in SO that relaxes the standard assumptions to study how the dome pattern may change under more realistic settings (e.g., [83, 86]). Tsai et al. [160] reformulate the original optimization model of the surgical scheduling problem as a two-stage mixed-integer model. The authors derive that the optimal solution of the second-stage problem is a boundary solution, which lays a good foundation for implementing SO.

In a stochastic simulation, specifying proper input distributions is often challenging and meanwhile, simulation cannot exactly evaluate the objective function due to randomness in the output. Empirical and statistical methods provide a way of gaining knowledge and information from the collected data. These methods can be used to identify/verify relationships or infer cause-and-effect relations for the phenomenon of interest in healthcare. Therefore, empirical and statistical methods could complement SO as follows.

First, they can provide a reliable estimation of the input parameters (e.g., [90, 181, 115, 182]) and validate the model assumptions (e.g.,[83, 48, 129]) for sophisticated SO models in healthcare. Kucukyazici et al. [90] integrate an empirical study on patient preferences in cancer screening programs with SO, which leads to a better estimation of the participation level and workload at each care facility and provides insights on the design of the cancer screening facility network in general. Marla et al. [115] use a rich dataset to model spatiotemporal distributions of emergency calls and to empirically estimate the ambulance abandonment behavior. The outcomes are incorporated into the SO model, which results in a reliable ambulance allocation. Klassen and Yoogalingam [83] and Fan et al. [48] assume lognormal service times for outpatient services as inputs since this distribution has been found empirically. Outside the reviewed literature, empirical/statistical methods are also widely used in this respect (e.g., [80, 146, 70]).

Second, they can provide a robust optimum of the SO for healthcare planning problems [165, 48, 77, 118, 119], and provide managerial insights based on the obtained solutions [57, 86]. Fan et al. [48] develop a robust selection procedure that selects the best alternative with some statistical guarantee under input uncertainty. Wang et al. [165] develop an SO method to sequentially construct a patient schedule using the statistical method of R&S, by which the best schedule is selected with high confidence. Furthermore, Goienetxea Uriarte et al. [57] combine data mining techniques (including statistical methods) and SO for ED capacity dimensioning. This provides decision makers not only with optimal solutions obtained by SO, but also with valuable knowledge about the variables and their interactions, which can be found in those solutions and will lead to the best capacity dimensioning. The latter is interesting since it can increase the practitioners' understanding of the solutions and increase the relevance of SO research to healthcare practice. As Terwiesch et al. [158] noted that one of empirical research's aims is to test operations management theory, empirical studies can be used to test whether the outcome of an SO model is supported by empirical evidence in healthcare.

8.3 Accurate simulation modeling for healthcare

The application of SO to healthcare is contingent on accurate simulation models for the healthcare systems of interest. On the one hand, if the simulation model is not an accurate representation of the system, any results derived from the model could be erroneous and may lead to costly decisions being made. Even, errors in healthcare simulation modeling might ultimately result in lives lost and the associated liabilities surrounding such events. Thus, the tolerable margin for error in healthcare simulation models is limited. On the other hand, seeking optimal values with inaccurate simulation models would be meaningless. Therefore, accurate modeling of healthcare systems requires significant attention, especially in the application of SO to healthcare. Section 8.2 has provided some important insights into accurate modeling. This section will highlight that not only model validation should be carefully performed, but also a suitable choice of the type of simulation plays a role.

There have been consistently ongoing efforts to develop accurate simulation models in healthcare, such as on hospital inpatient operations [146] and on hospital surgical department operations [140]. Validation against real data shows that their model reproduces key performance measures with some statistical guarantee. Several important aspects need attention when validating the healthcare simulation model. First, due to the complexity of healthcare simulation, model validation is not only something to be attempted after the model has already been developed, but also during each possible step of the model development. Also, we should validate not only the output of the overall simulation model, but also the model components (e.g., input distributions). Second, a better method for comparing system and model output data is to use historical system input data (if complete) to "drive" the model rather than samples from theoretical probability distributions. This (called the correlated inspection approach) could facilitate a statistically more precise comparison. Third, care should be taken for the statistical procedures for model validation (e.g., hypothesis testing and confidence intervals) since they require independent observations of the output performance measures. We recommend to use a confidence interval rather than a hypothesis test for model validation.

There are other useful validation techniques that are seldom used in the reviewed papers. First, one technique is formed by time-series approaches (a time series is a finite realization of a stochastic process) [96]. They are attractive as only one set of output data is needed from the system and one set from the model. Second, a good validation idea is to use one set of data for calibration during the process of improving the model and another independent set for final model validation [178]. Third, many papers aim to match the mean of the system data rather than the variance. However, it might be better to consider both the mean and the variance in certain healthcare planning situations. For example, when the aim is to manage and reduce variability in a healthcare system for improving performance, variation is a key part in simulation. In this case, the data distributions of the system and the model should sufficiently match. Otherwise it may seriously bias the results. For a comprehensive coverage on model validation, we refer to Sargent [143] and Law [96].

A suitable choice of the type of simulation is also important for accurate healthcare modeling, which depends on the nature of the underlying healthcare problem. There are mainly four simulation techniques applied in healthcare modeling, including DES, MCS, ABS and system dynamics (SD) [120, 139]. DES is a widely used simulation approach in healthcare operations and resource planning problems [139], while ABS is less frequently applied due to its relatively late emergence [120]. This is consistent with our findings. Compared to ABS (a bottom-up approach), SD is often used to model health systems from a more integrated (e.g., a hospital as a whole system) or top-level (e.g., healthcare policy evaluation) angle [120, 139]. In contrast, none of the reviewed works uses SD to model the healthcare facility. This constitutes a direction for future research.

On top of single-simulation models, there is an increasing interest in hybrid simulation (defined as models that combine at least two simulation approaches) to take advantage of the combined benefit [23]. Brailsford et al. [23] indicate that healthcare is the most frequent and potential area of application for hybrid simulation. However, the validation of hybrid models requires even more attention and effort than that of a single model since different types of simulation choices involve different validation methods and data requirements.

These advances in healthcare simulation can be brought into the SO paradigm for decision making. Attention should be paid to identifying certain structures inside specific simulation models in order to enhance the performance of SO. For example, ABS has a multi-level structure and often has a parameter-rich nature to characterize the behavior and the features of each agent. Exploiting the structural properties of the simulation model and the dependencies between parameters can reduce the search space size and thus, reduce the computational cost of parameter evaluation when integrating SO algorithms. When hybrid simulation is integrated into the SO paradigm, it is beneficial to choose appropriate simulation techniques at different stages of the entire SO process in order to reasonably use the time budget.

8.4 Promising SO algorithms for healthcare

With accurate simulation models, appropriate SO algorithms may be used to solve the healthcare planning problems. For comprehensive reviews about the recent development of SO, readers can refer to Amaran et al. [5] and Fu and Henderson [54].

A latest review on R&S is performed by Hong et al. [69]. The authors discuss recent achievements in adapting classical R&S to various practical problems, e.g., (1) large-scale R&S (2) constrained R&S, (3) multi-objective R&S, and (4) R&S with input uncertainty. These achievements could enhance R&S's potential to solve real-life healthcare planning problems. For ordinal optimization, Fu and Henderson [54] discuss that it has two key ideas that continue to heavily affect SO work today, including (1) it is easier to identify which of two solutions is better than to identify how much better, and (2) one might seek a near-optimal solution rather than a full-blown optimal solution. This might also be the case in healthcare applications. Healthcare practitioners might be satisfied with only one/two correct digits, because this is good enough for the application.

Direct search methods appear to be a promising option due to their simplicity, flexibility and reliability [89]. Although classical DSMs have been developed heuristically, Audet's survey [10] shows that recent DSMs have started to develop a convergence analysis. In addition, they can now handle multiple objectives, general constraints, and integer and categorical variables. These extend the ability of DSMs to solve more complex healthcare planning problems. In contrast to DSMs, gradient-based methods rely on gradient information. Although in principle it is often possible to obtain gradient information, general healthcare planning problems in reality can still make such attempts difficult.

The metamodel-based methods have a large appeal compared to other SO methods due to two distinct benefits [17]: (1) they obtain a deterministic metamodel response rather than a stochastic one, and (2) they result in shorter run times in general than the original simulation. Metamodel-based methods have proven to be beneficial in solving practical problems [54, 37], which might make their way into the healthcare domain in the coming years. For recent reviews, see Barton [16] and Kleijnen [87].

Random search methods are applicable to a broad range of problems [7]. They are also at the core of the SO software that is now widely used. Compared to the above SO algorithms, model-based methods use a probability distribution over the solution space (rather than any current set of solutions as in random search methods) to provide an estimate of more promising search areas [5]. They seem promising in SO to efficiently search for the solution.

In short, if the research is methodology-focused, a wide range of healthcare toy problems are available for demonstrating the performance of the proposed SO techniques. If the research is domain-focused or implementation-oriented in healthcare, the choice of SO algorithms might play a role. Those algorithms with a stronger ability to effectively account for real-life healthcare settings and to efficiently solve planning problems are promising. We also see great potential for many SO algorithms to leverage large-scale parallel computing that is now widely accessible [54], especially for those computationally-intensive algorithms.

9 Conclusion

This paper provides a comprehensive review of simulation optimization approaches for healthcare resource planning problems. SO has been shown to be effective for solving practical healthcare problems that, without too many simplifying assumptions, are too complex for exact/analytical techniques. We have demonstrated that this technique has become increasingly popular in recent years and in various care service planning problems. We have proposed a logical and structural classification method to categorize the articles.

By comparing the literature along different angles, we unveil the degree to which SO has been applied as well as trends that have arisen in this research field, as follows:

- The SO approach is most frequently applied to emergency care services and to strategic decisions (e.g., capacity dimensioning and ambulance location decisions).
- Patient access times to healthcare services, resource overtime, and financial value are the most used objectives. Multiple objectives are often considered and they are commonly combined into a single weighted objective function.
- DES is a popular simulation modeling method. Among the SO algorithms, random search methods (especially genetic algorithms) are the most popular. Metamodel-based methods (especially RSM and ANNs) take the second spot. In addition, optimization packages (especially OptQuest) are sometimes used.
- Most of the papers test the developed model with real-life data. However, SO appears to suffer from the same issues as other healthcare OR/MS approaches, i.e., lack of model validation of simulation and lack of implementation in practice.

- We also examine the connections between the individual classification fields by considering multiple dimensions simultaneously. Many intersections between various healthcare problem settings and SO approaches have not been researched to a large extent.

Based on our analysis of the literature, we identified major observations, insights and conclusions that deserve special attention when studying this field.

- *Where to use SO for healthcare.* SO is less widely used in healthcare resource planning than other OR/MS approaches. SO is suitable for incorporating complex and detailed problem settings, for evaluating objective functions without analytical forms, as well as for solving large problem instances in healthcare. However, the time, effort and data involved in developing, validating and experimenting the model are non-trivial.
- *How other methods complement SO for healthcare.* Effective healthcare resource planning might require an integration of SO with other OR/MS approaches. See Section 8.2 for how these methods could complement SO for healthcare planning.
- *Accurate simulation modeling for healthcare.* An accurate modeling of healthcare systems requires significant attention. For this, not only model validation should be carefully performed, but also a suitable choice of the type of simulation plays a role.
- *Promising SO algorithms for healthcare.* Various SO algorithms have progressed considerably and many of them have not widely made their way into the healthcare domain. We see great potential for many cutting-edge SO algorithms to be applied to healthcare planning, especially with the help of large-scale parallel computing.

We identify avenues for further research that provide opportunities for expanding existing methodologies and for narrowing the gap between theory and practice.

- *New algorithms for old problems.* Many healthcare resource planning problems have not yet fully been explored by SO, e.g., operating room capacity, home care service and residential care service planning. Moreover, for several already studied problems by SO, a comprehensive comparison study among different algorithms might be interesting.
- *New algorithms for new problems.* Real-life healthcare problems appear to be diverse regarding the type of decision variables, multiple objectives, uncertain parameters (input distributions and stochastic constraints), and other complex factors (e.g., behavioral factors and data availability). Thus, more advanced algorithms need to be developed, e.g., MOSO algorithms to identify Pareto solutions, robust SO and data-driven algorithms.

- *Impact on practice.* To increase the impact of the research on practice, it is necessary to see closer collaborations between researchers and practitioners in health systems. One key point is that the proposed method should be easy-to-implement (and -understand) on the practitioner side. Furthemore, the provision of key lessons on the actual implementation is beneficial to our research community.

Acknowledgements

This work was supported by the China Scholarship Council (No.201806890017). We would like to acknowledge the China Scholarship Council for the financial support to the first author.

References

[1] Aboueljinane, L., E. Sahin, Z. Jemai, and J. Marty (2014). A simulation study to improve the performance of an emergency medical service: Application to the french val-de-marne department. *Simulation Modelling Practice and Theory 47*, 46–59.

[2] Aghaabdellahian, Z. and M. Bijari (2020). Bed management considering bed-blocking and elective patient admissions using simulation optimisation. *International Journal of Simulation and Process Modelling 15*(3), 278–294.

[3] Ahmed, M. A. and T. M. Alkhamis (2009). Simulation optimization for an emergency department healthcare unit in kuwait. *European Journal of Operational Research 198*(3), 936–942.

[4] Allihaibi, W. G., M. Masoud, M. Elhenawy, S. Q. Liu, J. Burke, and A. Karim (2021). Solving the emergency care patient pathway by a new integrated simulation-optimisation approach. *IEEE Access 9*, 100895–100910.

[5] Amaran, S., N. V. Sahinidis, B. Sharda, and S. J. Bury (2016). Simulation optimization: A review of algorithms and applications. *Annals of Operations Research 240*(1), 351–380.

[6] Amorim, M., F. Antunes, S. Ferreira, and A. Couto (2019). An integrated approach for strategic and tactical decisions for the emergency medical service: Exploring optimization and metamodel-based simulation for vehicle location. *Computers & Industrial Engineering 137*, 106057.

[7] Andradóttir, S. (2015). A review of random search methods. *Handbook of Simulation Optimization*, 277–292.

[8] Anvaryazdi, S. F., S. Venkatachalam, and R. B. Chinnam (2020). Appointment scheduling at outpatient clinics using two-stage stochastic programming approach. *IEEE Access 8*, 175297–175305.

[9] Arnaout, J. P. (2010). Heuristics for the maximization of operating rooms utilization using simulation. *Simulation 86*(8-9), 573–583.

[10] Audet, C. (2014). A survey on direct search methods for blackbox optimization and their applications. *Mathematics Without Boundaries*, 31–56.

[11] Azcarate, C., L. Esparza, and F. Mallor (2020). The problem of the last bed: Contextualization and a new simulation framework for analyzing physician decisions. *Omega 96*, 102120.

[12] Azizi, M. J., F. Seifi, and S. Moghadam (2021). A robust simulation optimization algorithm using kriging and particle swarm optimization: Application to surgery room optimization. *Communications in Statistics-Simulation and Computation 50*(7), 2025–2041.

[13] Bae, K. H., M. Jones, G. Evans, and D. Antimisiaris (2019). Simulation modelling of patient flow and capacity planning for regional long-term care needs: A case study. *Health Systems 8*(1), 1–16.

[14] Baker, R. D. and P. L. Atherill (2002). Improving appointment scheduling for medical screening. *IMA Journal of Management Mathematics 13*(4), 225–243.

[15] Barton, R. R. (2009). Simulation optimization using metamodels. In *Proceedings of the 2009 Winter Simulation Conference (WSC)*, pp. 230–238. IEEE.

[16] Barton, R. R. (2020). Tutorial: Metamodeling for simulation. In *2020 Winter Simulation Conference (WSC)*, pp. 1102–1116. IEEE.

[17] Barton, R. R. and M. Meckesheimer (2006). Metamodel-based simulation optimization. In *Handbooks in Operations Research and Management Science*, Volume 13, pp. 535–574. Elsevier.

[18] Bazirha, M., A. Kadrani, and R. Benmansour (2021). Stochastic home health care routing and scheduling problem with multiple synchronized services. *Annals of Operations Research*, 1–29.

[19] Bekki, J. M., J. W. Fowler, G. T. Mackulak, and B. L. Nelson (2007). Using quantiles in ranking and selection procedures. In *2007 Winter Simulation Conference*, pp. 1722–1728. IEEE.

[20] Bendavid, I., Y. N. Marmor, and B. Shnits (2018). Developing an optimal appointment scheduling for systems with rigid standby time under pre-determined quality of service. *Flexible Services and Manufacturing Journal 30*(1-2), 54–77.

[21] Bertsimas, D., O. Nohadani, and K. M. Teo (2010). Robust optimization for unconstrained simulation-based problems. *Operations Research 58*(1), 161–178.

[22] Brailsford, S. C., M. W. Carter, and S. H. Jacobson (2017). Five decades of healthcare simulation. In *2017 Winter Simulation Conference (WSC)*, pp. 365–384. IEEE.

[23] Brailsford, S. C., T. Eldabi, M. Kunc, N. Mustafee, and A. F. Osorio (2019). Hybrid simulation modelling in operational research: A state-of-the-art review. *European Journal of Operational Research 278*(3), 721–737.

[24] Brailsford, S. C., P. R. Harper, B. Patel, and M. Pitt (2009). An analysis of the academic literature on simulation and modelling in health care. *Journal of Simulation 3*(3), 130–140.

[25] Braune, R., W. J. Gutjahr, and P. Vogl (2021). Stochastic radiotherapy appointment scheduling. *Central European Journal of Operations Research*, 1–39.

[26] Bélanger, V., E. Lanzarone, V. Nicoletta, A. Ruiz, and P. Soriano (2020). A recursive simulation-optimization framework for the ambulance location and dispatching problem. *European Journal of Operational Research 286*(2), 713–725.

[27] Cayirli, T. and E. D. Gunes (2014). Outpatient appointment scheduling in presence of seasonal walk-ins. *Journal of the Operational Research Society 65*(4), 512–531.

[28] Chan, C. W., V. F. Farias, N. Bambos, and G. J. Escobar (2012). Optimizing intensive care unit discharge decisions with patient readmissions. *Operations Research 60*(6), 1323–1341.

[29] Chen, C.-H., E. Yücesan, L. Dai, and H.-C. Chen (2009). Optimal budget allocation for discrete-event simulation experiments. *IIE Transactions 42*(1), 60–70.

[30] Chen, P.-S. and K.-L. Juan (2013). Applying simulation optimization for solving a collaborative patient-referring mechanism problem. *Journal of Industrial and Production Engineering 30*(6), 405–413.

[31] Chen, P.-S. and M.-H. Lin (2017). Development of simulation optimization methods for solving patient referral problems in the hospital-collaboration environment. *Journal of Biomedical Informatics 73*, 148–158.

[32] Chen, P. S., R. A. C. Robielos, P. Palana, P. L. L. Valencia, and G. Y. H. Chen (2015). Scheduling patients' appointments: Allocation of healthcare service using simulation optimization. *Journal of Healthcare Engineering 6*(2), 259–280.

[33] Chen, P.-S., K.-H. Yang, R. A. C. Robielos, R. A. C. Cancino, and L. A. M. Dizon (2016). Patient referral mechanisms by using simulation optimization. *Simulation Modelling Practice and Theory 61*, 14–27.

[34] Chen, T. L. and C. C. Wang (2016). Multi-objective simulation optimization for medical capacity allocation in emergency department. *Journal of Simulation 10*(1), 50–68.

[35] Chen, W., H. Guo, and K. L. Tsui (2020). A new medical staff allocation via simulation optimisation for an emergency department in Hong Kong. *International Journal of Production Research 58*(19), 6004–6023.

[36] Cildoz, M., A. Ibarra, and F. Mallor (2019). Accumulating priority queues versus pure priority queues for managing patients in emergency departments. *Operations Research for Health Care 23*, 100224.

[37] Coelho, G. F. and L. R. Pinto (2018). Kriging-based simulation optimization: An emergency medical system application. *Journal of the Operational Research Society 69*(12), 2006–2020.

[38] De Angelis, V., G. Felici, and P. Impelluso (2003). Integrating simulation and optimisation in health care centre management. *European Journal of Operational Research 150*(1), 101–114.

[39] De Santis, A., T. Giovannelli, S. Lucidi, M. Messedaglia, and M. Roma (2021). Determining the optimal piecewise constant approximation for the nonhomogeneous poisson process rate of emergency department patient arrivals. *Flexible Services and Manufacturing Journal*, 1–34.

[40] Deceuninck, M., S. De Vuyst, and D. Fiems (2019). An efficient control variate method for appointment scheduling with patient unpunctuality. *Simulation Modelling Practice and Theory 90*, 116–129.

[41] Dellino, G., J. P. Kleijnen, and C. Meloni (2012). Robust optimization in simulation: Taguchi and krige combined. *INFORMS Journal on Computing 24*(3), 471–484.

[42] Diefenbach, M. and E. Kozan (2011). Effects of bed configurations at a hospital emergency department. *Journal of simulation 5*(1), 44–57.

[43] d'Obrenan, A. V., A. Ridder, D. Roubos, and L. Stougie (2020). Minimizing bed occupancy variance by scheduling patients under uncertainty. *European Journal of Operational Research 286*(1), 336–349.

[44] Dogru, A. K. and S. H. Melouk (2019). Adaptive appointment scheduling for patient-centered medical homes. *Omega 85*, 166–181.

[45] Dong, J. and O. Perry (2020). Queueing models for patient-flow dynamics in inpatient wards. *Operations Research 68*(1), 250–275.

[46] Durand, G. A. and J. A. Bandoni (2020). Simulation-based optimization for the scheduling of elective surgery under uncertainty and downstream capacity constraints. *Journal of Systems Science and Systems Engineering 50*(2), 127–132.

[47] Erekat, A., G. Servis, S. C. Madathil, and M. T. Khasawneh (2020). Efficient operating room planning using an ensemble learning approach to predict surgery cancellations. *IISE Transactions on Healthcare Systems Engineering 10*(1), 18–32.

[48] Fan, W. W., L. J. Hong, and X. W. Zhang (2020). Distributionally robust selection of the best. *Management Science 66*(1), 190–208.

[49] Fan, X., J. Tang, C. Yan, H. Guo, and Z. Cao (2019). Outpatient appointment scheduling problem considering patient selection behavior: Data modeling and simulation optimization. *Journal of Combinatorial Optimization*, 1–23.

[50] Feng, Y. Y., I. C. Wu, and T. L. Chen (2017). Stochastic resource allocation in emergency departments with a multi-objective simulation optimization algorithm. *Health Care Management Science 20*(1), 55–75.

[51] Figueira, G. and B. Almada-Lobo (2014). Hybrid simulation–optimization methods: A taxonomy and discussion. *Simulation Modelling Practice and Theory 46*, 118–134.

[52] Fu, M. C. (2006). Gradient estimation. In *Handbooks in Operations Research and Management Science*, Volume 13, pp. 575–616. Elsevier.

[53] Fu, M. C., F. W. Glover, and J. April (2005). Simulation optimization: A review, new developments, and applications. In *Proceedings of the Winter Simulation Conference, 2005.*, pp. 13–pp. IEEE.

[54] Fu, M. C. and S. G. Henderson (2017). History of seeking better solutions, aka simulation optimization. In *2017 Winter Simulation Conference (WSC)*, pp. 131–157. IEEE.

[55] Ghanes, K., O. Jouini, A. Diakogiannis, M. Wargon, Z. Jemai, R. Hellmann, V. Thomas, and G. Koole (2015). Simulation-based optimization of staffing levels in an emergency department. *Simulation 91*(10), 942–953.

[56] Ghobadi, M., J. Arkat, H. Farughi, and R. Tavakkoli-Moghaddam (2021). Integration of facility location and hypercube queuing models in emergency medical systems. *Journal of Systems Science and Systems Engineering 30*(4), 495–516.

[57] Goienetxea Uriarte, A., E. Ruiz Zúñiga, M. Urenda Moris, and A. H. C. Ng (2017). How can decision makers be supported in the improvement of an emergency department? a simulation, optimization and data mining approach. *Operations Research for Health Care 15*, 102–122.

[58] Gokalp, E. (2020). Robust capacity planning for accident and emergency services. *RAIRO - Operations Research 54*(6), 1757–1773.

[59] Gonsalves, T. and K. Itoh (2009). Service optimization with patient satisfaction in healthcare systems. *Journal of Simulation 3*(3), 150–162.

[60] Granja, C., B. Almada-Lobo, F. Janela, J. Seabra, and A. Mendes (2014). An optimization based on simulation approach to the patient admission scheduling problem using a linear programing algorithm. *Journal of Biomedical Informatics 52*, 427–437.

[61] Grot, M., T. Becker, P. M. Steenweg, and B. Werners (2021). Enhanced coverage by integrating site interdependencies in capacitated ems location models. *Health Care Management Science*, 1–21.

[62] Guo, H., S. Gao, K. L. Tsui, and T. Niu (2017). Simulation optimization for medical staff configuration at emergency department in Hong Kong. *IEEE Transactions on Automation Science and Engineering 14*(4), 1655–1665.

[63] Guo, H., D. Goldsman, K.-L. Tsui, Y. Zhou, and S.-Y. Wong (2016). Using simulation and optimisation to characterise durations of emergency department service times with incomplete data. *International Journal of Production Research 54*(21), 6494–6511.

[64] Guo, H., H. Gu, Y. Zhou, and J. Peng (2020). A data-driven multi-fidelity simulation optimization for medical staff configuration at an emergency department in Hong Kong. *Flexible Services and Manufacturing Journal*, 1–25.

[65] Gupta, A., G. W. Evans, and S. S. Heragu (2013). Simulation and optimization modeling for drive-through mass vaccination - a generalized approach. *Simulation Modelling Practice and Theory 37*, 99–106.

[66] Hans, E. W., M. Van Houdenhoven, and P. J. H. Hulshof (2012). A framework for healthcare planning and control. In *Handbook of Healthcare System Scheduling*, pp. 303–320. Springer.

[67] Harper, P. R., N. H. Powell, and J. E. Williams (2010). Modelling the size and skill-mix of hospital nursing teams. *Journal of the Operational Research Society 61*(5), 768–779.

[68] Ho, Y.-C. (1999). An explanation of ordinal optimization: Soft computing for hard problems. *Information Sciences 113*(3-4), 169–192.

[69] Hong, L. J., W. Fan, and J. Luo (2020). Review on ranking and selection: A new perspective. *arXiv preprint arXiv:2008.00249*.

[70] Hu, W., C. W. Chan, J. R. Zubizarreta, and G. J. Escobar (2018). An examination of early transfers to the icu based on a physiologic risk score. *Manufacturing & Service Operations Management 20*(3), 531–549.

[71] Huang, X. M. (1998). Decision making support in reshaping hospital medical services. *Health Care Management Science 1*(2), 165–173.

[72] Huang, Y. L. (2016). The development of patient scheduling groups for an effective appointment system. *Applied Clinical Informatics 7*(01), 43–58.

[73] Hulshof, P. J., N. Kortbeek, R. J. Boucherie, E. W. Hans, and P. J. Bakker (2012). Taxonomic classification of planning decisions in health care: A structured review of the state of the art in OR/MS. *Health Systems 1*(2), 129–175.

[74] Hunter, S. R., E. A. Applegate, V. Arora, B. Chong, K. Cooper, O. Rincón-Guevara, and C. Vivas-Valencia (2019). An introduction to multiobjective simulation optimization. *ACM Transactions on Modeling and Computer Simulation (TOMACS) 29*(1), 1–36.

[75] INFORMS (2022). INFORMS: Operations Research in Healthcare. https://www.informs.org/Explore/History-of-O.R.-Excellence/O.R.-Application-Areas/Health-Care Accessed September 2022.

[76] Jagtenberg, C. J. and A. J. Mason (2020). Improving fairness in ambulance planning by time sharing. *European Journal of Operational Research 280*(3), 1095–1107.

[77] Kasaie, P. and W. D. Kelton (2013). Simulation optimization for allocation of epidemic-control resources. *IISE Transactions on Healthcare Systems Engineering 3*(2), 78–93.

[78] Kim, S. H. and Y. H. Lee (2016). Iterative optimization algorithm with parameter estimation for the ambulance location problem. *Health Care Management Science 19*(4), 362–382.

[79] Kim, S.-H. and B. L. Nelson (2006). Selecting the best system. In *Handbooks in Operations Research and Management Science*, Volume 13, pp. 501–534. Elsevier.

[80] Kim, S.-H. and W. Whitt (2014). Are call center and hospital arrivals well modeled by nonhomogeneous poisson processes? *Manufacturing & Service Operations Management 16*(3), 464–480.

[81] Kim, S.-H., W. Whitt, and W. C. Cha (2018). A data-driven model of an appointment-generated arrival process at an outpatient clinic. *INFORMS Journal on Computing 30*(1), 181–199.

[82] Klassen, K. J. and R. Yoogalingam (2008). An assessment of the interruption level of doctors in outpatient appointment scheduling. *Operations Management Research 1*(2), 95–102.

[83] Klassen, K. J. and R. Yoogalingam (2009). Improving performance in outpatient appointment services with a simulation optimization approach. *Production and Operations Management 18*(4), 447–458.

[84] Klassen, K. J. and R. Yoogalingam (2013). Appointment system design with interruptions and physician lateness. *International Journal of Operations and Production Management 33*(4), 394–414.

[85] Klassen, K. J. and R. Yoogalingam (2014). Strategies for appointment policy design with patient unpunctuality. *Decision Sciences 45*(5), 881–911.

[86] Klassen, K. J. and R. Yoogalingam (2019). Appointment scheduling in multi-stage outpatient clinics. *Health Care Management Science 22*(2), 229–244.

[87] Kleijnen, J. P. (2017). Regression and kriging metamodels with their experimental designs in simulation: A review. *European Journal of Operational Research 256*(1), 1–16.

[88] Kokangul, A., S. Akcan, and M. Narli (2017). Optimizing nurse capacity in a teaching hospital neonatal intensive care unit. *Health Care Management Science 20*(2), 276–285.

[89] Kolda, T. G., R. M. Lewis, and V. Torczon (2003). Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review 45*(3), 385–482.

[90] Kucukyazici, B., Y. Zhang, A. Ardestani-Jaafari, and L. Song (2020). Incorporating patient preferences in the design and operation of cancer screening facility networks. *European Journal of Operational Research 287*(2), 616–632.

[91] Kuo, Y. H., O. Rado, B. Lupia, J. M. Y. Leung, and C. A. Graham (2016). Improving the efficiency of a hospital emergency department: A simulation study with indirectly imputed service-time distributions. *Flexible Services and Manufacturing Journal 28*(1-2), 120–147.

[92] Lamiri, M., F. Grimaud, and X. Xie (2009). Optimization methods for a stochastic surgery planning problem. *International Journal of Production Economics 120*(2), 400–410.

[93] Lamiri, M., X. Xie, A. Dolgui, and F. Grimaud (2008). A stochastic model for operating room planning with elective and emergency demand for surgery. *European Journal of Operational Research 185*(3), 1026–1037.

[94] Landa, P., R. Aringhieri, P. Soriano, E. Tànfani, and A. Testi (2016). A hybrid optimization algorithm for surgeries scheduling. *Operations Research for Health Care 8*, 103–114.

[95] Landa, P., M. Sonnessa, E. Tànfani, and A. Testi (2018). Multiobjective bed management considering emergency and elective patient flows. *International Transactions in Operational Research 25*(1), 91–110.

[96] Law, A. M. (2015). *Simulation modeling and analysis* (5th Edition ed.). New York: McGraw-Hill Education.

[97] Lee, E. K., H. Y. Atallah, M. D. Wright, E. T. Post, C. I. Thomas, D. T. Wu, and J. Haley, L. L. (2015). Transforming hospital emergency department workflow and patient care. *Interfaces 45*(1), 58–82.

[98] Lee, E. K., S. Maheshwary, J. Mason, and W. Glisson (2006). Decision support system for mass dispensing of medications for infectious disease outbreaks and bioterrorist attacks. *Annals of Operations Research 148*(1), 25–53.

[99] Lee, E. K., F. Pietz, B. Benecke, J. Mason, and G. Burel (2013). Advancing public health and medical preparedness with operations research. *Interfaces 43*(1), 79–98.

[100] Lee, M. L., I. Park, D. U. Park, and C. Park (2017). Constrained ranking and selection for operations of an emergency department. *International Journal of Simulation Modelling 16*(4), 563–575.

[101] Li, N., X. Li, C. Zhang, and N. Kong (2021). Integrated optimization of appointment allocation and access prioritization in patient-centred outpatient scheduling. *Computers & Industrial Engineering 154*, 107125.

[102] Li, N., J. Pan, and X. Q. Xie (2020). Operational decision making for a referral coordination alliance - when should patients be referred and where should they be referred to? *Omega 96*, 102077.

[103] Li, N., Y. Zhang, D. Teng, and N. Kong (2021). Pareto optimization for control agreement in patient referral coordination. *Omega 101*, 102234.

[104] Li, S., G. Koole, and X. Xie (2020). An adaptive priority policy for radiotherapy scheduling. *Flexible Services and Manufacturing Journal 32*(1), 154–180.

[105] Liang, F., Y. Y. Guo, and R. Y. K. Fung (2015). Simulation-based optimization for surgery scheduling in operation theatre management using response surface method. *Journal of Medical Systems 39*(11).

[106] Lin, C. K. Y., T. W. C. Ling, and W. K. Yeung (2017). Resource allocation and outpatient appointment scheduling using simulation optimization. *Journal of Healthcare Engineering 2017*.

[107] Lin, R. C., M. Y. Sir, and K. S. Pasupathy (2013). Multi-objective simulation optimization using data envelopment analysis and genetic algorithm: Specific application to determining optimal resource levels in surgical services. *Omega 41*(5), 881–892.

[108] Liu, R., X. Xie, K. Yu, and Q. Hu (2018). A survey on simulation optimization for the manufacturing system operation. *International Journal of Modelling and Simulation 38*(2), 116–127.

[109] Liu, Z., D. Rexachs, F. Epelde, and E. Luque (2017). A simulation and optimization based method for calibrating agent-based emergency department models under data scarcity. *Computers & Industrial Engineering 103*, 300–309.

[110] Lucidi, S., M. Maurici, L. Paulon, F. Rinaldi, and M. Roma (2016a). A derivative-free approach for a simulation-based optimization problem in healthcare. *Optimization Letters 10*(2), 219–235.

[111] Lucidi, S., M. Maurici, L. Paulon, F. Rinaldi, and M. Roma (2016b). A simulation-based multiobjective optimization approach for health care service management. *IEEE Transactions on Automation Science and Engineering 13*(4), 1480–1491.

[112] Luo, Y. and E. Lim (2013). Simulation-based optimization over discrete sets with noisy constraints. *IISE Transactions 45*(7), 699–715.

[113] Mallor, F. and C. Azcárate (2014). Combining optimization with simulation to obtain credible models for intensive care units. *Annals of Operations Research 221*(1), 255–271.

[114] Mallor, F., C. Azcárate, and J. Barado (2016). Control problems and management policies in health systems: Application to intensive care units. *Flexible Services and Manufacturing Journal 28*(1-2), 62–89.

[115] Marla, L., K. Krishnan, and S. Deo (2021). Managing EMS systems with user abandonment in emerging economies. *IISE Transactions 53*(4), 389–406.

[116] McCormack, R. and G. Coates (2015). A simulation model to enable the optimization of ambulance fleet allocation and base station location for increased patient survival. *European Journal of Operational Research 247*(1), 294–309.

[117] Miranda, D. M. M. and M. M. P. M. Miranda (2021). Discrete-event simulation applied to a radiotherapy process: A case study of a cancer center. *Brazilian Journal of Operations & Production Management 18*(1), 1–20.

[118] Molina-Pariente, J. M., E. W. Hans, and J. M. Framinan (2018). A stochastic approach for solving the operating room scheduling problem. *Flexible Services and Manufacturing Journal 30*(1-2), 224–251.

[119] Mousavi, H., S. A. Darestani, and P. Azimi (2021). An artificial neural network based mathematical model for a stochastic health care facility location problem. *Health Care Management Science*, 1–16.

[120] Mustafee, N., K. Katsaliaki, and S. J. Taylor (2010). Profiling literature in healthcare simulation. *Simulation 86*(8-9), 543–558.

[121] Nguyen, L., E. Bellucci, and L. T. Nguyen (2014). Electronic health records implementation: An evaluation of information system impact and contingency factors. *International Journal of Medical Informatics 83*(11), 779–796.

[122] Niessner, H., M. S. Rauner, and W. J. Gutjahr (2018). A dynamic simulation–optimization approach for managing mass casualty incidents. *Operations Research for Health Care 17*, 82–100.

[123] Niu, Q., Q. Peng, and T. Y. ElMekkawy (2013). Improvement in the operating room efficiency using tabu search in simulation. *Business Process Management Journal 19*(5), 799–818.

[124] OECD (2022). Data from the website of Organisation for Economic Co-operation and Development (OECD). http://www.oecd.org/health/ Accessed September 2022.

[125] Olafsson, S. and J. Kim (2002). Simulation optimization. In *Proceedings of the Winter Simulation Conference, 2002*, Volume 1, pp. 79–84. IEEE.

[126] Oliveira, B., J. A. de Vasconcelos, J. F. F. Almeida, and L. R. Pinto (2020). A simulation-optimisation approach for hospital beds allocation. *International Journal of Medical Informatics 141*, 104174.

[127] Oueida, S., Y. Kotb, S. Ionescu, and G. Militaru (2019). Ams: A new platform for system design and simulation. *International Journal of Simulation Modelling 18*(1), 33–46.

[128] Ozcan, Y. A., E. Tànfani, and A. Testi (2017). Improving the performance of surgery-based clinical pathways: A simulation-optimization approach. *Health Care Management Science 20*(1).

[129] Pan, X., N. Geng, and X. Xie (2021). Appointment scheduling and real-time sequencing strategies for patient unpunctuality. *European Journal of Operational Research*.

[130] Paul, J. A. and R. Batta (2011). Improving hurricane disaster preparedness: Models for optimal reallocation of hospital capacity. *International Journal of Operational Research 10*(2), 194–213.

[131] Peng, Y., X. Qu, and J. Shi (2014). A hybrid simulation and genetic algorithm approach to determine the optimal scheduling templates for open access clinics admitting walk-in patients. *Computers & Industrial Engineering 72*(1), 282–296.

[132] Petering, M. E. H., O. T. Aydas, K. Kuzu, and A. Ross (2015). Simulation analysis of hospital intensive care unit reimbursement policies from the triple bottom line perspective. *Journal of Simulation 9*(2), 86–98.

[133] Porter, M. E. et al. (2010). What is value in health care. *New England Journal of Medicine 363*(26), 2477–2481.

[134] Qiu, Y., J. Song, and Z. Liu (2016). A simulation optimisation on the hierarchical health care delivery system patient flow based on multi-fidelity models. *International Journal of Production Research 54*(21), 6478–6493.

[135] Rabbani, M., A. Farshbaf-Geranmayeh, and R. Yazdanparast (2018). A simulation optimization approach for integrated resource allocation in an emergency department, pharmacy, and lab. *Intelligent Decision Technologies 12*(2), 187–212.

[136] Rezaeiahari, M. and M. T. Khasawneh (2020). Simulation optimization approach for patient scheduling at destination medical centers. *Expert Systems with Applications 140*, 112881.

[137] Rohleder, T., B. Bailey, B. Crum, T. Faber, B. Johnson, L. Montgomery, and R. Pringnitz (2013). Improving a patient appointment call center at mayo clinic. *International Journal of Health Care Quality Assurance 26*(8), 714.

[138] Romero, H. L., N. P. Dellaert, S. van der Geer, M. Frunt, M. H. Jansen-Vullers, and G. A. M. Krekels (2013). Admission and capacity planning for the implementation of one-stop-shop in skin cancer treatment using simulation-based optimization. *Health Care Management Science 16*(1), 75–86.

[139] Roy, S., S. Prasanna Venkatesan, and M. Goh (2020). Healthcare services: A systematic review of patient-centric logistics issues using simulation. *Journal of the Operational Research Society*, 1–23.

[140] Samudra, M., E. Demeulemeester, B. Cardoen, N. Vansteenkiste, and F. E. Rademakers (2017). Due time driven surgery scheduling. *Health Care Management Science 20*(3), 326–352.

[141] Saremi, A., P. Jula, T. ElMekkawy, and G. G. Wang (2013). Appointment scheduling of outpatient surgical services in a multistage operating room department. *International Journal of Production Economics 141*(2), 646–658.

[142] Saremi, A., P. Jula, T. ElMekkawy, and G. G. Wang (2015). Bi-criteria appointment scheduling of patients with heterogeneous service sequences. *Expert Systems with Applications 42*(8), 4029–4041.

[143] Sargent, R. G. (2013). Verification and validation of simulation models. *Journal of Simulation 7*(1), 12–24.

[144] Sasanfar, S., M. Bagherpour, and A. Moatari-Kazerouni (2020). Improving emergency departments: Simulation-based optimization of patients waiting time and staff allocation in an iranian hospital. *International Journal of Healthcare Management*, 1–8.

[145] Schütz, H. J. and R. Kolisch (2012). Approximate dynamic programming for capacity allocation in the service industry. *European Journal of Operational Research 218*(1), 239–250.

[146] Shi, P., M. C. Chou, J. G. Dai, D. Ding, and J. Sim (2016). Models and insights for hospital inpatient operations: Time-dependent ed boarding time. *Management Science 62*(1), 1–28.

[147] Shi, P., J. E. Helm, J. Deglise-Hawkinson, and J. Pan (2021). Timing it right: Balancing inpatient congestion vs. readmission risk at discharge. *Operations Research 69*(6), 1842–1865.

[148] Shi, Y., T. Boudouh, and O. Grunder (2017). A hybrid genetic algorithm for a home health care routing problem with time window and fuzzy demand. *Expert Systems with Applications 72*, 160–176.

[149] Shirazi, B. (2016). Fast track system optimization of emergency departments: Insights from a computer simulation study. *International Journal of Modeling, Simulation, and Scientific Computing 7*(03), 1650015.

[150] Siler, K. F. (1979). Evaluation of emergency ambulance characteristics under several criteria. *Health Services Research 14*(2), 160.

[151] Sinreich, D. and O. Jabali (2007). Staggered work shifts: A way to downsize and restructure an emergency department workforce yet maintain current operational performance. *Health Care Management Science 10*(3), 293–308.

[152] Sinreich, D., O. Jabali, and N. P. Dellaert (2012). Reducing emergency department waiting times by adjusting work shifts considering patient visits to multiple care providers. *IISE Transactions 44*(3), 163–180.

[153] Song, J., Y. Qiu, and Z. Liu (2016). Integrating optimal simulation budget allocation and genetic algorithm to find the approximate pareto patient flow distribution. *IEEE Transactions on Automation Science and Engineering 13*(1), 149–159.

[154] Sun, Y. and X. P. Li (2013). Response surface optimisation of surgery start times in a single operating room using designed simulation experiments. *International Journal of Healthcare Technology and Management 14*(1-2), 61–72.

[155] Sundaramoorthi, D., V. C. P. Chen, J. M. Rosenberger, S. B. Kim, and D. F. Buckley-Behan (2010). A data-integrated simulation-based optimization for assigning nurses to patient admissions. *Health Care Management Science 13*(3), 210–221.

[156] Swoveland, C., D. Uyeno, I. Vertinsky, and R. Vickson (1973). A simulation-based methodology for optimization of ambulance service policies. *Socio-Economic Planning Sciences 7*(6), 697–703.

[157] Tellis, R., O. Starobinets, M. Prokle, U. N. Raghavan, C. Hall, T. Chugh, E. Koker, S. C. Chaduvula, C. Wald, and S. Flacke (2021). Identifying areas for operational improvement and growth in ir workflow using workflow modeling, simulation, and optimization techniques. *Journal of Digital Imaging 34*(1), 75–84.

[158] Terwiesch, C., M. Olivares, B. R. Staats, and V. Gaur (2020). Om forum—a review of empirical operations management over the last two decades. *Manufacturing & Service Operations Management 22*(4), 656–668.

[159] Troy, P., N. Lahrichi, D. Porubska, and L. Rosenberg (2020). Fine-grained simulation optimization for the design and operations of a multi-activity clinic. *Flexible Services and Manufacturing Journal 32*(3), 599–628.

[160] Tsai, S. C., Y. Yeh, and C. Y. Kuo (2021). Efficient optimization algorithms for surgical scheduling under uncertainty. *European Journal of Operational Research 293*(2), 579–593.

[161] Uribe-Sánchez, A. and A. Savachkin (2010). Two resource distribution strategies for dynamic mitigation of influenza pandemics. *Journal of Multidisciplinary Healthcare 3*, 65–77.

[162] Uribe-Sánchez, A., A. Savachkin, A. Santana, D. Prieto-Santa, and T. K. Das (2011). A predictive decision-aid methodology for dynamic mitigation of influenza pandemics. *OR Spectrum 33*(3), 751–786.

[163] Vancroonenburg, W., P. De Causmaecker, and G. Vanden Berghe (2019). Chance-constrained admission scheduling of elective surgical patients in a dynamic, uncertain setting. *Operations Research for Health Care 22*, 18.

[164] Visintin, F., P. Cappanera, C. Banditori, and P. Danese (2017). Development and implementation of an operating room scheduling tool: An action research study. *Production Planning & Control 28*(9), 758–775.

[165] Wang, D., D. J. Morrice, K. Muthuraman, J. F. Bard, L. K. Leykum, and S. H. Noorily (2018). Coordinated scheduling for a multi-server network in outpatient pre-operative care. *Production and Operations Management 27*(3), 458–479.

[166] Wang, X., X. Gong, N. Geng, Z. Jiang, and L. Zhou (2020). Metamodel-based simulation optimisation for bed allocation. *International Journal of Production Research 58*(20), 6315–6335.

[167] Wing, J. and P. Vanberkel (2021). Simulation optimisation for mixing scheduled and walk-in patients. *Health Systems*, 1–12.

[168] Woodall, J. C., T. Gosselin, A. Boswell, M. Murr, and B. T. Denton (2013). Improving patient access to chemotherapy treatment at duke cancer institute. *Interfaces 43*(5), 449–461.

[169] Xiao, Y. and R. Yoogalingam (2021). Reserved capacity policies for operating room scheduling. *Operations Management Research 14*, 107–122.

[170] Xie, J. and P. I. Frazier (2013). Sequential bayes-optimal policies for multiple comparisons with a known standard. *Operations Research 61*(5), 1174–1189.

[171] Xie, T., M. Ni, Z. Zhang, and Y. Wei (2019). Parallel simulation decision-making method for a response to unconventional public health emergencies based on the scenario–response paradigm and discrete event system theory. *Disaster Medicine and Public Health Preparedness 13*(5-6), 1017–1027.

[172] Yao, H. C., P. J. Chen, Y. T. Kuo, C. C. Shih, X. Y. Wang, and P. S. Chen (2020). Solving patient referral problems by using bat algorithm. *Technology and Health Care 28*(S1), 433–442.

[173] Yeh, J. and W. Lin (2007). Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department. *Expert Systems with Applications 32*(4), 1073–1083.

[174] Yip, K., L. Leung, and D. Yeung (2018). Levelling bed occupancy: Reconfiguring surgery schedules via simulation. *International Journal of Health Care Quality Assurance 31*(7), 864–876.

[175] Yousefi, M., M. Yousefi, R. P. M. Ferreira, J. H. Kim, and F. S. Fogliatto (2018). Chaotic genetic algorithm and adaboost ensemble metamodeling approach for optimum resource planning in emergency departments. *Artificial Intelligence in Medicine 84*, 23–33.

[176] Yu, G., A. Liu, and H. Sun (2021). Risk-averse flexible policy on ambulance allocation in humanitarian operations under uncertainty. *International Journal of Production Research 59*(9), 2588–2610.

[177] Zaffar, M. A., H. K. Rajagopalan, C. Saydam, M. Mayorga, and E. Sharer (2016). Coverage, survivability or response time: A comparative study of performance statistics used in ambulance location models via simulation–optimization. *Operations Research for Health Care 11*, 1–12.

[178] Zeinali, F., M. Mahootchi, and M. M. Sepehri (2015). Resource planning in the emergency departments: A simulation-based metamodeling approach. *Simulation Modelling Practice and Theory 53*, 123–138.

[179] Zhang, H., T. J. Best, A. Chivu, and D. O. Meltzer (2020). Simulation-based optimization to improve hospital patient assignment to physicians and clinical units. *Health Care Management Science 23*(1), 117–141.

[180] Zhang, Y. and M. L. Puterman (2013). Developing an adaptive policy for long-term care capacity planning. *Health Care Management Science 16*(3), 271–279.

[181] Zhang, Y., M. L. Puterman, M. Nelson, and D. Atkins (2012). A simulation optimization approach to long-term care capacity planning. *Operations Research 60*(2), 249–261.

[182] Zhang, Z., B. P. Berg, B. T. Denton, and X. Xie (2019). Appointment scheduling and the effects of customer congestion on service. *IISE Transactions 51*(10), 1075–1090.

[183] Zhang, Z. and X. L. Xie (2015). Simulation-based optimization for surgery appointment scheduling of multiple operating rooms. *IISE Transactions 47*(9), 998–1012.

[184] Zhen, L., K. Wang, H. Hu, and D. Chang (2014). A simulation optimization framework for ambulance deployment and relocation problems. *Computers & Industrial Engineering 72*, 12–23.

[185] Zhou, L. P., N. Geng, Z. B. Jiang, and X. X. Wang (2018). Multi-objective capacity allocation of hospital wards combining revenue and equity. *Omega 81*, 220–233.