Planning and Scheduling in Healthcare for Better Care Coordination: Current Understanding, Trending Topics, and Future Opportunities

Seokjun Youn
Eller College of Management
The University of Arizona
Tucson, AZ 85721
Email: syoun@arizona.edu
Phone: (520) 626-0493

H. Neil Geismar
Mays Business School
Texas A&M University
College Station, TX 77843
Email: ngeismar@mays.tamu.edu
Phone: (979) 458-0033

Michael Pinedo*
Stern School of Business
New York University
New York City, NY 10012
Email: mpinedo@stern.nyu.edu
Phone: (212) 998-0287

* *Corresponding Author*

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/poms.13867

Planning and Scheduling in Healthcare for Better Care Coordination: Current Understanding, Trending Topics, and Future Opportunities

Seokjun Youn
Department of Management Information Systems, Eller College of Management, The University of Arizona, 1130 E Helen St., McClelland Hall 430CC, Tucson, AZ 85721, USA, syoun@arizona.edu

H. Neil Geismar
Department of Information and Operations Management, Mays Business School, Texas A&M University, 4217 TAMU, Wehner 320F, College Station, TX 77843, USA, ngeismar@mays.tamu.edu

Michael Pinedo*
Department of Information, Operations, and Management Sciences, Stern School of Business, New York University, 44 West 4th Street, 8-65. New York City, New York 10012, USA, mpinedo@stern.nyu.edu

This study reviews the research on planning and scheduling in healthcare published in *Production and Operations Management (POM)* and other prestigious operations management journals over the past 30 years, identifies trending topics, and discusses opportunities for future research. Specifically, we categorize the capacity planning literature into three problem contexts: a) hospital units and operating rooms, b) outpatient clinics, and c) broader healthcare networks. We further classify the scheduling literature into a) appointment scheduling, b) surgery and workforce scheduling, and c) recurring and integrated scheduling, with several subtopics that have attracted attention in the literature. Afterward, we demonstrate the seven emerging trends in healthcare (i.e., collaborative ecosystem and cooperative competition strategy, expanded reach of virtual care, applications of artificial intelligence in healthcare, patient engagement and personalized care, shifting to value-based care, working toward health equity, and aging populations) and list relevant research opportunities and their related departments of *POM* journal. We conclude with a call for multidisciplinary perspectives to deliver coherent solutions that incorporate the complex contemporary issues harmoniously.

Key words: Healthcare; Capacity Planning; Scheduling; Literature Review; Production and Operations Management
History: Received: March 2022; Accepted April 2022 by Christopher S. Tang and Subodha Kumar.

1. Introduction

Healthcare planning, scheduling, and control assist in determining *what*, *when*, *where*, and *how much* to deliver, which are important issues in healthcare operations management research. Modern healthcare systems are under pressure to improve efficiency, to enhance quality of care, to widen access, and to reduce inequity. Achieving these conflicting goals requires innovation, new technologies, and improved processes (Keskinocak and Savva 2020). Lack of strategic and operational planning in healthcare organizations often results in service disconnection and in reduced value to the patient per effort by the provider. Healthcare productivity relies largely on the appropriate planning and scheduling of all activities involved, and the range of planning and scheduling processes is enormous (Pinedo 2005).

Operations management (OM) scholars have responded by adapting our field's planning and scheduling techniques to the decisions faced by stakeholders in healthcare systems, including patients, medical professionals, healthcare organizations, and policymakers. This article focuses on planning and scheduling issues in healthcare to celebrate the insights gained by *Production and Operations Management (POM)* since its inception 30 years ago and to reflect on trending topics and future research opportunities.

Research included in this review was published in *POM* (between 1992 and 2022) and in other prestigious operations management journals, including *Management Science (ManSci)*, *Manufacturing & Service Operations Management (M&SOM)*, and *Operations Research (OperRes)*. From the journals' online libraries, we collected the initial set of 280 papers published in *POM* and 322 papers published in either *ManSci, M&SOM*, or *OperRes*, using one of or combinations of the following keywords: "healthcare," "scheduling," "capacity," "planning," "patient," and "physician." We excluded any publication that was not related to healthcare operations management or that identified healthcare as a potential application while primarily focusing on a different area (e.g., manufacturing). This procedure identified the 59 *POM*, 12 *ManSci*, 19 *M&SOM*, and 12 *OperRes* papers cited in this study. Our review is necessarily brief because of the page limit imposed by this special issue, so only representative examples are covered in each area. We were unable to include numerous other excellent papers in this review; any errors or oversights are our own.

In this study, we highlight the major findings and implications of each discussed work. These papers apply various methodologies, such as mathematical programming, analytical models, queuing theory, simulation, econometrics, and experiments, to problems relevant to planning and scheduling in healthcare. In §2, we begin with a discussion of the capacity management literature and further divide our review along three major problem contexts, i.e., hospital units, outpatient clinics, and broader healthcare networks. Most research on specific hospital units focuses on operating rooms, for which resources are expensive and critical to other perioperative processes. More general reviews on operating room management can be found in Cayirli and Veral (2003), Gupta (2007), Gupta and Denton (2008), Cardoen et al. (2010), May et al. (2011), and Zhu et al. (2019). In §3, we review how the scheduling literature in healthcare has evolved over the past decades to tackle many interrelated operational uncertainties for patients, care providers, and administrators. For ease of reference, we categorize the scheduling literature into appointment scheduling, surgery and workforce scheduling, and recurring and integrated scheduling, with several subtopics. However, these categorizations are not mutually exclusive because many studies deal with problems that encompass two or more areas. §4 identifies emerging trends in healthcare and discusses several opportunities for future research. The emerging trends are motivated by the advancement of information/clinical technologies, lessons from the pandemic, evolving research methods, and changing patient needs that directly or indirectly influence our current understanding of capacity planning and scheduling. Therefore, these call for healthcare operations management researchers' attention for continuing improvement in this vital industry.

2. Capacity Management in Healthcare

Capacity management in healthcare aims for the efficient use of internal resources to meet fluctuating demand for services (Smith-Daniels et al. 1988, Gupta et al. 2016). The term *capacity* generally refers to any measure of a system's processing ability enabled by its resources (Van Mieghem 2003), including physical space, equipment, and workforce. Capacity management and scheduling are inseparably related to each other, although capacity decisions are strategic and scheduling ones are operational (Zhu et al. 2019). As an earlier work in this domain, Roth and Van Dierdonck (1995) attempt to apply knowledge accumulated in manufacturing to healthcare settings. Specifically, they develop a hospital operations planning and control system based on diagnostic-related groups (DRGs) that define health services as *products* and on manufacturing resources planning (MRP-II).

In what follows, we discuss the evolution of capacity management studies by operations management scholars on various hospital units, including operating rooms (§2.1), outpatient clinics (§2.2), and broader healthcare networks (§2.3).

2.1. Hospital Units and Operating Rooms

Hospitals have multiple functional units, such as an emergency department (ED), intensive care units (ICUs), operating rooms (ORs), and inpatient wards. One third of the U.S. healthcare expenditures occur in hospitals (Martin et al. 2019), and ORs are generally the largest cost centers as well as revenue sources among all hospital units (Childers and Maggard-Gibbons 2018, Jung et al. 2019). Naturally, much research considers OR capacity management.

When a hospital anticipates a surge in surgical cases, new OR capacity can be achieved by either building new ORs or extending the hours of operation in existing ORs. Lovejoy and Li (2002) look into the trade-offs between three performance metrics: patient wait times, scheduled procedure start-time reliability, and hospital revenues. Surgery allocation to ORs is a difficult combinatorial optimization problem. The fact that surgical procedures have such a wide range of durations further complicates assignment decisions. Denton et al. (2010) presents stochastic optimization models for a particular day's surgery assignment to ORs. They consider both the fixed cost of opening ORs and the variable cost of overtime relative to a certain length of day in the objective function.

Hospitals have limited capacity not only for operating rooms, but also for inpatient beds, when demand spikes. This lowers the quality of patient care by prolonging the time between admission and placement on a floor. Furthermore, the inability to provide assistance to potential patients results in income loss. In the event of a demand spike, Thompson et al. (2009) suggest proactively shifting patients between floors. In the quest of optimal capacity allocation, units are frequently examined jointly. Patients going through critical care and acute care, or acute care and post-acute care, for example, are considered by Bretthauer et al. (2011). If the second stage is crowded, a patient who has completed their first stage service will be barred from exiting the first stage. The authors devise a heuristic for tandem systems that allows them to quickly assess the effects of blocking on system performance. Step down units (SDUs), which provide an intermediate level of care between ICUs and standard medical-surgical wards, may be used to avoid blocking. Armony et al. (2018) proposes a patient flow queueing model in the ICU and SDU to identify when an SDU is needed, what size it should be, and what factors influence these decisions.

Hospital administrators frequently seek to make the most of a limited number of beds by providing services for a variety of care types. To solve this problem, Best et al. (2015) investigate the partitioning of care types among a hospital's specialized units, develop an optimization model for specialized unit formation decision, and address the benefits of devoted services endogenously. The administrator must choose between large units, which pool demand and bed capacity, and specialized units, which each specialize in a specific sort of care types. Beds for high-utility care types can also be reserved in specialized units. The authors discover that as the hospital is divided into specialized units, overall bed occupancy declines. Shorter periods of stay coupled with targeted care, on the other hand, may boost overall patient throughput if appropriate focus is achieved. They have also noticed that when people are ready to wait longer for admission, the hospital tends to create extra units. As a result, hospitals with higher wait times can create more specialized units and so provide more concentrated care.

More recently, Bavafa et al. (2019) consider the challenge of allocating daily hospital service capacity among many types of elective surgical procedures in the presence of urgent procedures that could interrupt the elective procedures. Many demand types compete for multiple types of service capacity in the resulting model, which is a multidimensional form of the inverse newsvendor problem. In the presence of urgent procedures, the authors show how to put together the best portfolios of elective procedures. Given the wide range of levels of urgency in high-volume operating rooms, capacity allocation decisions can have a significant impact on how wait times are rationed. Carew et al. (2021) looks at a longer-term sequential capacity planning problem in which a hospital assigns operating room time to several surgical specialties. The objective is to minimize an urgency-weighted wait time measure. Their solution improves the number of patients treated within their urgency-based wait time constraints by up to 21% by defining the capacity allocation problem as a Markov Decision Process (MDP) and employing simulation-based dynamic programming to overcome a huge state space. In a similar vein, Mills et al. (2021) study operational strategies that improve surge capacity and identify how these can be most effectively deployed based on the characteristics of individual hospitals.

A comprehensive effort by Naderi et al. (2021) studies a generic operating room planning and scheduling problem in which patients and resources (i.e., operating rooms, surgeons, anesthesiologists) are assigned to days, resources are assigned to patients, and patients are sequenced in each day. The objective is to develop a weekly operation plan that reduces fixed and overtime expenses. They discovered that Toronto General Hospital in Ontario, Canada, can increase surgery loads by 40% using existing OR safety capacities. They also make suggestions for how the hospital might alter its downstream capacities to accommodate varied degrees of surgical volume growth. This can be used to the urgent requirement for more capacity that has arisen as a result of the COVID-19 pandemic.

2.2. Outpatient Clinics

Though primary-care and outpatient clinics often face challenges in capacity management that differ from those of hospitals, their capacity issues are still intertwined with scheduling concerns, such as same-day or future appointments. For example, primary-care clinics suffer uncertainty from patient choices in addition to variable patient arrivals. Patients have varying judgments of the intensity of their need, as well as diverse time-of-day preferences and levels of commitment to their primary-care practitioner. Although enhanced access technologies are intended to shorten wait times and improve patient satisfaction by allowing patients to select their own appointment times, the clinic must carefully regulate patient access to physician appointments. Too many appointments scheduled in advance might lead to capacity issues and longer wait times for patients. Conversely, scheduling too few appointments increases the likelihood of clinic slots going underused. The capacity management issue here is deciding which appointment requests to accept in order to maximize income. Gupta and Wang (2008) develops an MDP model for this appointment-booking problem.

Because capacity, patient flow, and scheduling are rarely handled together in outpatient clinics, researchers wonder whether clinic performance can be enhanced if the policies that guide these decisions are set together. White et al. (2011) develop an empirically based discrete-event simulation to investigate the interactions between appointment scheduling policies and capacity allocation policies (i.e., the number of available examination rooms) and to see how they interact to affect various performance measures, including resource utilization and patient waiting time. The results suggest that scheduling lower-variance, shorter appointments earlier leads in less overall patient waiting time without affecting physician utilization or lengthening clinic visits.

A clinic's planning decisions may incorporate other factors, such as the level of patients' treatment adherence. For infectious diseases such as tuberculosis and HIV, treatment adherence is critical in treatment efficiency and epidemic management. While studies of some infectious diseases show that patients who reside closer to their health facilities have better adherence, most models overlook patient heterogeneity. McCoy and Johnson (2014) develop an optimization model that integrates a clinic's capacity decisions with population health outcomes. In a case study of the HIV epidemic in Zambia, they find that incorporating adherence into clinic planning models can lead to decisions that significantly improve outcomes.

Specialty clinics care for patients referred by primary care physicians, emergency departments, or other specialists. Urgent patients are frequently seen on the referral day, but non-urgent referrals are usually scheduled for a later date. Clinics must determine how much appointment capacity is required to achieve reasonably quick access for non-urgent patients and to produce a balanced performance. To help discover the capacity that leads to the desired performance, Izady (2015) examine the dynamics of an appointment backlog as a discrete-time bulk service queue and offers numerical methods for efficient computation of relevant performance measures. Several practical characteristics are captured in the models, including arbitrary referral and appointment cancellation distributions, delay-dependent no-show behavior, and no-show rescheduling. The author shows how the models can accurately forecast performance and how they may be used to design appointment capacity.

The conventional first-come, first-served strategy to scheduling outpatient appointments ignores different levels of urgency, resulting in needlessly long wait times for urgent patients. Deglise-Hawkinson et al. (2018) develop a capacity allocation optimization methodology that reserves appointment slots for urgent situations in a complex, integrated care setting where many specialists serve multiple sorts of patients to address this issue. This optimization reallocates network capacity to reduce access delays (i.e., indirect waiting times) for initial and downstream appointments that are prioritized according to urgency. To concurrently smooth workloads and guarantee access latency targets, the authors define this problem as a queueing optimization and approximate it using deterministic linear optimization. A case study reveals the potential to minimize urgent patient mean access delay by 27% while only increasing non-urgent patient mean access delay by 7%. With the same service levels and overtime, throughput increases by 31%.

Even in clinics, patients may go though multiple stages, such as holding rooms, operating rooms, and post-anesthesia care units, in a single day. The interdependence of activities, their stochastic durations, and the uncertainties in patient-mix pose significant challenges in managing the capacity of each activity and in achieving a smooth patient flow by coordinating the stages for each patient's visit. Youn et al. (2022) develop a framework to plan capacity in each of three sequential stages for ambulatory surgery centers (ASCs). In contrast to the traditional top-down approach to capacity planning, their approach proposes a bottom-up strategy that uses optimization methods and data analytics based on operational-level patient flow information. They model ASCs as hybrid flow shops, then relax the fixed capacity assumption of traditional hybrid flow shop problems by using the trade-off between the overtime cost and the amortized capacity construction cost.

2.3. Broader Healthcare Networks

Several studies discuss the capacity planning of a healthcare system network for community care services (Kucukyazici et al. 2011, Bidhandi et al. 2019, Cherkesly et al. 2019), which are becoming increasingly important as patients live longer. Kucukyazici et al. (2011) provide a framework for physicians and system planners to construct chronic illness management strategies and community-based care designs. They use a patient flow approach to describe the multiple care-provider visits of patients with a specific chronic condition, which extends analytical epidemiologic models. A Markov model based on a disease-specific state space is used to capture the patterns of care received by a patient group in compact form. The patients' case-mix and care-provider clustering are reflected in their framework. The framework is used to analyze data from over 4,000+ stroke patients who were discharged from acute care hospitals in Quebec, Canada.

Bidhandi et al. (2019) propose a queuing network approach to network capacity planning. They optimize capacity allocation across the network using simulated annealing, with a performance guarantee based on the sum of the blocking probabilities. The approach is applied to a local health region with a network that includes acute care, long-term care, assisted living, home care, rehabilitation, and chronic care. Cherkesly et al. (2019) develop tools to aid in the creation and analysis of a community healthcare network in order to expand health coverage in underserved areas. The results of an application in Liberia are shown.

Gökalp et al. (2020) investigate a capacity planning issue in a network of stem-cell donation centers. The underlying optimization model integrates donor search procedures with the goal of increasing the number of transplants. The effect of demand and service time variabilities on the capacity planning framework is investigated using a scenario-based stochastic programming approach. To obtain resilient answers against uncertainties, the maximum possible waiting time throughout the search process is evaluated.

3. Scheduling in Healthcare: Managing Operational Uncertainty in Diverse Contexts and Scopes

In this section, we review the advances in the healthcare operations management literature on scheduling. The majority of studies use algorithmic decision support tools based on operations research techniques to use critical and costly resources efficiently. An emerging body of empirical and experimental research complements the literature to confirm and validate many contextual behavioral issues that are often assumed in such algorithmic decision support tools. Whereas the previous section on planning involves decisions on overall capacity to meet uncertain demand, scheduling deals with more granular decisions such as the types, number, and sequence of patients to be seen on a particular day or week at various types of healthcare facilities. We classify the large body of literature into the following three categories[1] and highlight the key lessons: i) appointment scheduling (§3.1) that often has to cope with patient no-shows, adherence, and overbooking, which are closely related to one another, ii) surgery and workforce scheduling (§3.2) including block scheduling with various objectives, and iii) recurring and integrated scheduling (§3.3) such as periodic scheduling of multiple encounters for a particular patient, possibly over different departments or facilities.

3.1. Appointment Scheduling

A critical contributing factor in optimizing patient appointment schedules is to comprehend various types of scheduling methods and to apply the best suited one to the problem context. The types of appointment scheduling include, but are not limited to, *time-slot scheduling* (i.e., first come first serve scheduling), *wave scheduling* (which brings several patients in at the same time, e.g., at the beginning of each hour instead of specifying times every 15 or 20 min during the hour), *open booking* (i.e., walk-in appointment scheduling), *priority scheduling* (where certain appointments are

[1] These categories are neither mutually exclusive nor collectively exhaustive.

reserved for urgent needs). The different types are often combined to compensate for their respective weaknesses in improving patient flow and in securing utilization levels under various uncontrollable factors such as patient no-show and demand surge. For example, open booking is one of the most flexible types because patients are allowed to attend an appointment anytime on a day that suits them best, but, naturally, controlling overcrowding becomes a significant challenge under a demand surge. Thus, a variation that combines the walk-in component with wave scheduling can be effective as it provides patients not only the flexibility of walk-in appointments but also the security of a designated appointment window. Healthcare OM scholars have researched the correct balance between the patients who booked appointments in advance and those who are waiting. In general, the information gathered from patient flow analysis and cycle-time management help to identify inefficiencies that result in longer patient visits that lead to overtime, increased patient wait time, decreased number of visits per day, increased no-show rates, and reduced revenue. In general, the literature collectively emphasizes the importance of clarifying problem contexts and incorporating the practitioners' perspectives.

3.1.1. Prioritizing Patients in Diagnostic and Primary Care Facilities: Earlier studies focused on understanding patient arrival dynamics at diagnostic and primary care facilities (Green et al. 2006, Patrick et al. 2008, Dobson et al. 2011). Hospital diagnostic facilities, such as magnetic resonance imaging centers, serve diverse patient groups, such as outpatients, who are scheduled in advance, inpatients, whose requests for service are produced at random intervals throughout the day, and emergency patients, who must be handled as soon as possible. Green et al. (2006) focus on developing the outpatient appointment schedule and establishing dynamic priority rules for admitting patients for the service given these distinct patient types. They use a finite-horizon dynamic program to model the problem of managing patient demand for diagnostic services. Their numerical investigations provide insight into the optimal policies' sensitivity to various cost and probability characteristics. They also assess the effectiveness of a number of heuristic rules for accepting appointments and scheduling patients.

Patrick et al. (2008) describe a method for scheduling patients with varying priority at a diagnostic facility in a public health-care system dynamically. The resource manager's challenge, rather than maximizing income, is to dynamically allocate available capacity to incoming demand in order to meet wait time targets in a cost-effective manner. They use an MDP to represent the scheduling process, use approximation dynamic programming to solve the equivalent linear program, and use simulation to test the policy's quality.

In a primary care institution, it is typical to set aside slots for urgent patients. The effect of this practice on two service quality measures is examined in Dobson et al. (2011): the average number of urgent patients not handled during normal hours (either handled as overtime, referred to other physicians, or referred to the emergency room) and the average queue of non-urgent or routine patients. They show how patient arrival dynamics affect performance quality and how encouraging routine patients to call for same-day appointments is a significant component of advanced-access success.

Variability in patient arrival and service times can result in long wait times for patients and inefficient use of facility resources. Salzarulo et al. (2011) empirically investigate how different sources of variability affect a primary care facility's performance. A physician's ancillary tasks, patient timeliness, unannounced visits to the facility's laboratory or X-ray services, brief interruptions of a patient's examination, and examination time variation by patient class are all examples of these sources. Unscheduled visits to the facility's laboratory or X-ray services have the most impact on a physician's idle time, according to their findings. How the physician prioritizes ancillary tasks, such as phone calls, against examining patients, has the greatest impact on the average patient wait time.

3.1.2. Patient Classification and Appointment Intervals: No two patients are the same, even if they come in with the same symptoms. Nevertheless, grouping patients into a smaller number of classes could be beneficial not only for building tractable models but also for applying the models in practice. In this regard, several studies explore appointment rules based on patient classification and appointment intervals (Cayirli et al. 2008, Klassen and Yoogalingam 2009, Cayirli et al. 2012, Samorani and Ganguly 2016). Cayirli et al. (2008) look into two different applications of patient classification: only for sequencing patient appointments and for both sequencing and appointment interval adjustment. Appointment intervals are adjusted in the latter technique to meet the consultation time characteristics of different patient classes. According to their simulation, the interval adjustments for patient classes are successful in improving doctors' idle time, doctors' overtime, and patients' waiting times without any trade-offs.

Klassen and Yoogalingam (2009) examine optimal rules for a stochastic appointment scheduling problem using simulation. The "dome" scheduling rule (in which appointment intervals are initially short, progressively grow toward the middle of the session, and then drop near the conclusion of the session) is found to be reliable, but practitioners may benefit from considering a flatter, "plateau-dome" scheduling rule. The plateau-dome scheduling pattern has been proved to be reliable across a wide range of performance metrics and circumstances. Cayirli et al. (2012) further investigate a universal dome appointment rule that can be parameterized through a planning constant that is characterized for each clinic by environmental factors such as no-shows, walk-ins, number of appointments per session, variability of service times, and cost of doctor's time vs. patients' time. They also discuss an appointment system adjustment mechanism to specifically reduce the disruption caused by no-shows and walk-ins.

Patients frequently arrive early and out of turn for scheduled appointments in outpatient clinics, prompting Samorani and Ganguly (2016) to study whether an available physician should see an early patient immediately away (preempt) or wait for the next patient scheduled. This "wait-preempt dilemma" is especially pertinent in high-service clinics like psychotherapy, chiropractic, or acupuncture, because preempting may result in a little late patient waiting for an unnecessarily long time. They analytically determine the time intervals for which provider preemption is optimal and those for which provider waiting is optimal, even in the presence of waiting patients.

3.1.3. Patient's Satisfaction, Preference, and Information: Unlike manufacturing processes, providers must consider satisfaction and preference of patients when scheduling them. OM scholars have incorporated such issues into their models from early 2010 (e.g., Wang and Gupta 2011, Salzarulo et al. 2011, Feldman et al. 2014, Salzarulo et al. 2016). Patients' satisfaction with a system for a non-urgent appointments is affected by their ability to choose a doctor and to select a convenient time of day. The difficulty of matching patients' booking requests with physicians' available slots in a way that improves both patient satisfaction and clinic revenue is a major barrier when creating outpatient appointment systems. The fact that booking preferences are not tracked, may differ from one patient to the next, and may alter over time exacerbates the problem. Wang and Gupta (2011) propose a design framework for appointment systems that dynamically learn and update patients' preferences and use this information to improve booking selections. Motivated by the growing use of electronic appointment booking systems, Feldman et al. (2014) develop appointment scheduling models that consider the patient time preferences. The service provider dynamically decides which appointment days to make available for the patients.

Other individual patient factors, such as return or new visit, health status, patient timeliness, supplementary tasks, are used by Salzarulo et al. (2011) to improve appointment scheduling performance. Patients are sequenced based on their classifications, which enhances system performance by up to 25.5%, according to the researchers. Salzarulo et al. (2016) further look into how patient data can be used to predict patient examination durations in the clinic's appointment scheduling system, in order to see if using individual patient characteristics is better than using a traditional classification method. Computational results show that this approach of patient scheduling can save up to 24.2% on an overall cost function that includes patient wait time, physician idle time, and overtime, especially when patients are sequenced with short-duration patients first.

Although service systems are stochastic, their appointment systems are often deterministic. Customer punctuality and service durations, for example, are frequently considered to be equal to their means throughout the design stage of healthcare services. Motivated by the gap between what is planned and reality, Mandelbaum et al. (2020) consider appointment scheduling and sequencing under a time-varying number of servers in a data-rich environment where service durations and punctuality are variable. Their data-driven method, which is built on infinite-server queues, results in solutions that are tractable and scalable. The authors show that their strategy regularly reduces waiting and overtime expenses by 15% to 40% in cancer center infusion units with around 90 daily visits and 25+ infusion chairs under a variety of experimental setups.

3.1.4. Patient No-shows: Missed Appointments, Missed Opportunities: No-shows cost the U.S. health care system more than $150 billion a year and individual physicians an average of

$200 per unused time slot (Forbes 2019). OM scholars have devoted extensive efforts to tackling the huge operational inefficiency caused by patient no-shows.

Clinics have a particularly serious no-show problem, with reported no-show rates ranging from 3% to 80%. No-shows reduce revenues and provider productivity, increase costs, and limit patient access by reducing effective clinic capacity. To mitigate the detrimental effects of patient no-shows, LaGanga and Lawrence (2012) develop a fast and effective solution procedure for constructing near-optimal overbooked appointment schedules by balancing the benefits of servicing additional patients with the possible costs of patient waiting and clinic overtime. Zacharias and Pinedo (2014) further investigate an overbooking model for scheduling arrivals at a medical institution in which patients have varying no-show probabilities and weights based on customer classes (e.g., diagnoses). The scheduler sends patients to time slots to reduce the projected weighted sum of patients' waiting times and doctors' idle time and overtime. After analyzing the static problem in which the set of patients and their characteristics are known in advance, they apply their findings to the sequential scheduling problem, in which requests for appointments come in gradually and the scheduler must assign each patient to one of the remaining slots for a given day. They find that the no-show rate and the heterogeneity of the patients have a substantial impact on the optimal schedule. Chen and Robinson (2014) study appointment scheduling for a mix of routine patients who made appointments ahead of time and last-minute patients who call for an appointment on the same day. They decide when to schedule these same-day patients throughout the day and how the possibility of their arrivals affects the routine patients' appointment timings. The optimal patient sequence is highly dependent on the likelihood of no-shows and the projected number of same-day patients.

While there are a variety of reasons why people fail to show up for appointments, empirical research have showed that the likelihood of a patient being a no-show grows as the time between the call for the appointment and the appointment date increases. Liu and Ziya (2014) and Liu (2016) each use queueing models to show that patient sensitivity to incremental delays plays a more important role than do the no-show probabilities in determining the optimal appointment window, i.e., the time frame in which patients are allowed to make appointments.

In outpatient clinics, Lee et al. (2018) study appointment block scheduling policies for single providers under both patient heterogeneity in service times and patient no-shows. The objective is to find daily appointment schedules that minimize the weighted sum of patient wait time, physician idle time, and physician overtime. Their sequential block scheduling procedures are motivated by the successful Toyota Production System load smoothing approach.

Overbooking and patient appointment reminders are the two main strategies clinics use to mitigate the negative effects of no-shows. Developing efficient overbooking schedules necessitates a patient-level estimate of communication sensitivity, whereas developing effective reminder systems requires a patient-level estimate of communication sensitivity. No such forecasts can be made using current approaches for calculating no-show rates at the patient level. To address this, Li et al.

(2019) develop a Bayesian nested logit model that estimates individual-level coefficients for patient-specific predictors using appointment confirmation data. When no-show odds are low, the value comes mostly from lowering waiting time; when no-show percentages are high, the benefit comes primarily from reducing physician overtime and idle time. This study claims that the Bayesian method enables customized appointment reminders, based on a patient's confirmation behavior, and improved overbooking scheduling, especially in clinics with a large patient throughput.

Kong et al. (2020) investigate how to schedule medical appointments in the face of time-dependent patient no-shows and unpredictable service durations. They use a distributionally robust model that optimizes patient scheduled arrival times to reduce the worst-case total estimated expenditures of patient waiting and service provider's idle and overtime. Taking into account the time-of-day fluctuation in patient no-show rates results in a considerable reduction in total estimated cost.

When short-term base appointment capacity is filled, the scheduler must choose between delaying an appointment and risking a costly failure, such as readmission, or scheduling the appointment sooner using surge capacity at a higher cost. The majority of appointment scheduling literature focuses on the trade-off between waiting times and utilization. In contrast, Grant et al. (2022) analyze preventative appointment scheduling and its impact on the broader service supply network when the service provider is responsible for service and failure costs. A stochastic dynamic program that includes no-shows characterizes the optimal stochastic scheduling policy and evaluates the performance of heuristics. Numerical experiments reveal that intuitive appointment policies utilized in practice are robust under modest capacity usage, but their optimality gap can quadruple under a large load.

3.2. Surgery and Workforce Scheduling

Surgical suites are a key driver of a hospital's expenditures, revenues, and postoperative resource use such as beds (Gupta 2007). In this section, we summarize surgery and workforce scheduling studies under several topics such as surgical block scheduling, patient waiting time versus surgical suite overtime, concurrent scheduling of elective and emergency patients, nursing scheduling, and scheduling that takes provider's characteristics into account.

3.2.1. Optimizing Surgical Block Scheduling: The procedure of allocating OR resources to a surgeon or group of surgeons for a specific day and time is known as block scheduling. Chow et al. (2011) describe a transparent and portable strategy that combines a Monte Carlo simulation with a mixed integer program (MIP) to optimize block scheduling procedures. For a specified surgery schedule, simulation samples from historical cases record and predict bed requirements, assuming no resource restrictions. To reduce peak bed occupancies, the MIP model supports the simulation model by scheduling both surgeon blocks and patient categories. Day et al. (2012) investigate the difficulty of balancing two competing goals in the pursuit of efficient operating room management in a hospital: providing surgeons with predictable, reliable access to operating rooms and maintaining high capacity utilization. The common solution in practice to the first problem is to grant exclusive block time by assigning an operating room to a particular surgeon for a portion of the week. As a significant advance over this existing approach, the authors model the prospect of "shared" block time, which satisfies capacity constraints only in expectation.

3.2.2. Patient Waiting Time versus Surgical Suite Overtime: Many additional research have attempted to reconcile the two conflicting objectives of minimizing patient waiting time and surgical suite overtime. We describe two representative ones. Gul et al. (2011) compare multiple scheduling heuristics for outpatient centers. A discrete event simulation model evaluates how 12 different sequencing and patient appointment time-setting heuristics perform with respect to the competing criteria and compared with current practice. After that, a bi-criteria genetic algorithm is used to see whether there are any better solutions for this single-day scheduling challenge. More recently, Zhou et al. (2021) consider a surgical scheduling problem with the objective of minimizing the expected makespan while staying within a pre-specified upper bound on each patient's expected waiting time. The authors adopt two approximation methods because of its complex combinatorial and nonlinear nature. In the certainty equivalent method, they approximate the expected time span of the cases by assuming that the waiting time of each surgical case is a deterministic value. In the variance bounding method, they approximate the expected time span based on upper bounds on the variance of the surgical case waiting times. They demonstrate that the confidence equivalent method gives a lower bound on the optimal expected time span, while the variance bounding method gives an upper bound. Both methods advise arranging surgical cases in ascending order of time variability as a simple sequencing rule.

3.2.3. A Long-standing Dilemma: Scheduling Elective and Emergency Patients Together: Hospital beds are a critical resource, so finding the right admission balance between elective and emergency patients has been widely discussed (Helm et al. 2011, Meng et al. 2015, Truong 2015, Freeman et al. 2016, Jung et al. 2019). The variability in hospital occupancy, if not adequately managed, has a detrimental impact on the cost and quality of patient treatment by increasing emergency department (ED) congestion, emergency blocks and diversions, elective cancellations, ancillary service backlogs, overstaffing, and understaffing. Controlling inpatient admissions can help to alleviate these issues by lowering variability in hospital occupancy.

The ED and scheduled elective admission are the two most common gateways to enter a hospital. Excessive wait times in highly utilized hospitals make the scheduled gateway undesirable or infeasible for a portion of patients. As a result, this group often seeks admission to the hospital through the ED. Helm et al. (2011) propose a third gateway, i.e., an expedited patient care queue, to better serve these patients and improve overall hospital operations. They use a Markov decision model to characterize an optimal admission threshold policy employing controls on the scheduled and expedited gateways.

Alternatively, enforcing quotas on elective admissions could alleviate shortfalls. Meng et al. (2015)

present a distributionally robust optimization approach for managing elective admissions to determine these quotas. Based on an ambiguous set of probability distributions, they propose an optimum budget of variation strategy that maximizes the level of uncertainty the admission system can withstand without violating the expected bed shortfall constraint. Meanwhile, dynamic assignment of appointments to manage daily variations in demand and capacity is especially challenging because of its high dimensionality. Truong (2015) tackles this issue by using a canonical model of dynamic scheduling with two patient classes: an urgent demand class that must be treated on the day of arrival, and a regular demand class that can be served at a later date. In their model, patients are presumed to take the first available appointment and make no distinction between providers.

Elective surgery scheduling has a substantial impact on a hospital's financial health and other metrics that are relevant to stakeholders. Freeman et al. (2016) develop a scheduling formulation that takes into account the uncertainty of elective surgery durations and plans for the possibility of unexpected urgent demands. Using a scenario-based modeling approach, the study indicates that incorporating uncertainty via scenarios boosts profit and OR utilization as compared to deterministic scheduling methods. Furthermore, explicitly considering urgent arrivals significantly reduces the time that patients of this type must wait for assistance, with little influence on other essential metrics. In the context of Level-1 trauma hospitals, Jung et al. (2019) address the issue of allocating limited capacity to emergent surgery cases while scheduling elective patients. Specifically, they develop a model for allocating the OR capacity to elective patients so that the emergency patients who arrive randomly can be served without experiencing long waits.

3.2.4. Nursing Shortage and Scheduling: The nursing profession continues to experience shortages because of high turnover, a scarcity of educators, and inequitable workforce distribution (Flinkman et al. 2010, Halter et al. 2017). In most acute care hospitals, the ongoing scarcity has a negative impact on productivity, quality of care, and operating expenses. Such nursing shortages have been addressed using scheduling approaches developed by OM researchers. Easton et al. (1992), for example, examined expected nursing expense and workforce requirements under 12 different scheduling approaches that are expected to improve turnover. Nurse absenteeism, which is common in U.S. hospitals, drives up inpatient staffing expenses. Wang and Gupta (2014) study which factors, including unit culture, short-term workload, and shift type, explain nurse absenteeism. The authors develop models to explore the impact of demand and absentee rate variability on staffing plan performance and come up with some structural findings. These findings are used to develop and test heuristics for identifying near-optimal staffing strategies.

Nurses can work in several areas of the ED, but their assignments are only altered at the beginning of their shifts. This prompted Chan et al. (2021) to investigate the dynamic assignment of servers to various areas of a service system at the beginning of discrete time intervals, or shifts. This partial flexibility provides an opportunity for reducing the expected waiting time of customers. They study a discrete-time fluid control problem to minimize transient holding costs over a finite horizon and find that introducing partial flexibility by reassigning servers at the beginning of shifts can substantially reduce the expected system cost by 10%–50% compared to using dedicated staffing.

3.2.5. Provider's Satisfaction and Productivity: Recently, several studies have started exploring providers' satisfaction/preference issues, such as rotation scheduling for medical training (Lemay et al. 2017, Cire et al. 2019) and incorporating physician productivity (Zaerpour et al. 2022). When it comes to scheduling medical residents, the objective is often to optimize resident satisfaction within the range of viable schedules, relative to the many hard constraints that ensure appropriate patient coverage and adequate training opportunities. The frequency of time-off requests granted is a common metric of resident satisfaction. However, simply maximizing this total may result in undesirable schedules because certain requests are more important than others. Instead, Lemay et al. (2017) propose compiling an entire list of maxially feasible and minimally infeasible sets of requests, which schedulers can use to select their preferred option. The proposed method could be used to solve a wide range of problems with soft restrictions. In the context of medical school students, Cire et al. (2019) investigate the scheduling practices that must assign a cohort of students to a series of clinical rotations of varying periods at hospitals in various locations while meeting operational and quality-of-service requirements.

Although the literature tackles stochastic patient arrivals and care durations, physician productivity is often assumed to be stable. Zaerpour et al. (2022) relax this assumption by assigning physicians to shifts so that ED wait times are reduced without adding new physicians. In particular, they extend the physician rostering problem by including heterogeneity across emergency physicians in terms of productivity, as measured by the number of new patients seen in one hour, as well as the stochastic nature of patient arrivals and physician productivity. A two-stage stochastic programming is used to model their physician rostering problem. A simulation study calibrated using real data suggests that the new scheduling method can lower patient wait times by as much as 13% compared to the ED's present scheduling system, which does not account for varying levels of physician productivity.

3.3. Recurring and Integrated Scheduling

Whereas many studies focus on a single visit of a particular patient in a particular department of a facility, the OM community has gradually expanded our scheduling knowledge base to patient types who need periodic care or who need multidisciplinary or multistage services. We highlight a few such studies.

3.3.1. Scheduling Patients with Periodic Visits: Patients with a chronic illness may require therapy on a regular basis. Yu et al. (2020) focus on appointment scheduling for healthcare institutions with patients who are booked for multiple appointments rather than just one; radiotherapy/chemotherapy for cancer, physical therapy, kidney dialysis, and diabetic treatment are examples. The appointment scheduling problem is modeled as an MDP model in which huge state space motivates an index policy based on a one-step policy improvement algorithm. The index policy provides a significant improvement over heuristics such as the Next Available Day Policy and the Shortest Queue Policy.

Patients with end-stage renal disease are commonly hospitalized for complications from their treatment, resulting in dialysis clinic capacity being unused. Overbooking at the clinic is appealing because of these sporadic patient absences. Lee and Zenios (2009) develop a migration network to capture patient flow into the clinic and between the clinic and the hospital. They find that High hospitalization rates and protracted inpatient stays allow for more overbooking. Even while keeping the probability of capacity shortage arbitrarily small, numerical examples based on a typical dialysis clinic in the U.S. suggest that policies that allow overbooking increase earnings by 11% -14% over policies derived from traditional M/M/N models that do not allow overbooking.

3.3.2. Multidisciplinary and Multistage Scheduling: Broadening the problem scope to multiple specialties and multiple stages substantially increases the problem's complexity. Nevertheless, decision-making from such an inclusive view helps avoid suboptimality and thereby improves the overall system efficiency.

Much attention has been paid to long wait times in EDs, and much research has sought to improve ED performance. However, ED congestion is often exacerbated by the inability to move patients into wards, which are sometimes clogged with patients waiting for a bed in a long-term care (LTC) facility. Patrick (2011) present an MDP model that determines the required number of available LTC beds to keep the census of patients waiting for LTC in hospitals below a certain threshold. Similarly, Price et al. (2011) investigate the problem of allocating surgical block times to days and surgical groups in order to balance the flow into and out of the intensive care unit (ICU). They provide evidence that the model achieves the goal of reducing overnight stays in post-anesthesia care units, which occur because of congestion in the ICU.

Controlling admissions across multiple resources is particularly challenging. Hospitals generally lack appropriate enterprise-level strategic planning of bed and care resources, resulting in statistically out-of-control census levels. Bed block, surgical cancellations, ambulance diversions, and operational chaos are all symptoms of system breakdown. To optimally solve the strategic planning and scheduling problem, Helm and Van Oyen (2014) develop analytical models of a controlled hospital census that are then incorporated into a MIP. Their solution method aligns elective admissions with other hospital subsystems to alleviate system congestion. This work provides the theoretical foundations for an efficient scheduled admissions planning system as well as a practical decision-making mechanism to keep hospital census stable. Barz and Rajaram (2015) investigate a patient admission problem in a hospital with numerous resource restrictions (e.g., ORs and beds) and a stochastic evolution of patient care requirements across multiple resources. For a random stream of non-emergency elective patients, the hospital must determine whether to accept, postpone, or even reject admission. They formulate the control process as an MDP with an objective of maximizing expected contribution net of overbooking costs, develop bounds using approximate dynamic programming, and use these bounds to build heuristics.

Various modeling and problem-solving techniques have been devised and applied to multispecialty and multistage contexts. Rath et al. (2017) examine the problem of minimizing daily expected resource usage and overtime costs across multiple parallel resources such as anesthesiologists and ORs at large multispecialty hospitals. They devise a two-stage mixed-integer stochastic dynamic program with recourse. They further develop a data-driven robust optimization method for solving large-scale real-sized versions of this model that are close to optimality. To trade off resource utilization and overtime expenses, this approach efficiently incorporates the flexibility of resources and the uncertainty of surgery lengths. This has resulted in an increase of 3.5% in anesthesiologist utilization and a 3.8% rise in operating room usage, as well as a daily cost reduction of roughly 7%. Diamant et al. (2018) further investigate the scheduling practices of a multidisciplinary, multi-stage, outpatient healthcare program. Patients undergo a battery of tests before being considered for elective surgery. High rates of attrition and appointment no-shows plague such systems, resulting in capacity underutilization and treatment delays. They propose a scheduling model in which the clinic assigns patients to an appointment day but waits to see who shows up before deciding which assessments they will receive. The clinic gains flexibility as a result, allowing it to increase system performance. They employ approximation dynamic programming to solve the scheduling problem, which they formulate as an MDP.

To optimize the mix of patients' medical conditions prior to surgery, Wang et al. (2018) develop a coordinated pre-operative scheduling approach between anesthesiology and internal medicine. The idea of combining these two services has conceptual appeal because any health issues detected by an anesthesiologist can usually be addressed by a general internist. The problem is formulated as a two-station stochastic network, in which each station, i.e., clinic, may have numerous service providers staffed in parallel, and patients visit the first available provider. The objective is to strike a balance between profit and patient wait times and clinic overtime costs. They develop a scheduling method with a booking limit to create a balanced network schedule and demonstrate the benefit of the proposed approach on a healthcare network managed by The University of Texas Health Sciences Center in San Antonio.

Despite the fact that hospital care is frequently administered in stages, current healthcare scheduling and capacity planning systems regard distinct hospital units usually as separate entities. To address this shortcoming, Liu et al. (2019b) propose an MDP model for scheduling surgical patients on a daily basis, explicitly taking into account both patient length-of-stay after surgery and inpatient census. They find that the conventional scheduling method, which is solely based on operating room utilization, can result in highly inefficient downstream capacity utilization and up to a three-fold increase in total expenses. In contrast, a scheduling strategy based on downstream capacity consumption, often performs similarly to the integrated scheduling policy, and so may serve as a simple, effective scheduling heuristic for hospital administrators.

Currently, outpatient programs require patients to make their own appointment decisions and coordinate their care. This approach may result in a rise in missed appointments and unreasonable access delays, among other inefficiencies. Wang et al. (2019) propose a multistation network model that balances assumptions that allow tractability and that prevent real-world adoption. To enable real-world applicability, they explore sequential appointment scheduling in a network of stations with exponential service delays, no-show possibilities, and overbooking. They present a heuristic myopic scheduling policy that is close to optimal. However, because it is not simple enough for practical implementation, they further explore a series of approximations and find one that delivers a significant computational advantage.

4. Emerging Trends in Healthcare and Future Research Directions

As we have discussed so far, healthcare OM scholars have devised ideas to improve operational efficiency and uncovered the impacts of numerous risk factors on planning/scheduling performance in healthcare. In Figure 1, we illustrate an overview of the strategic approach to planning and scheduling in healthcare in terms of objectives, attributes of a successful program, risk types, and a few steps to scaling planning/scheduling decisions. Exploring each or some of the components in this figure will most likely continue as our complex healthcare systems face new advanced technologies, evolving patient needs, and unexpected events over time. For example, the COVID-19 pandemic has revealed how vulnerable the healthcare industry is to change and its need for structural and technological transformation. Gupta et al. (2021) review pandemics/epidemics research published in major operations management, operations research, and management science journals and identify research gaps.

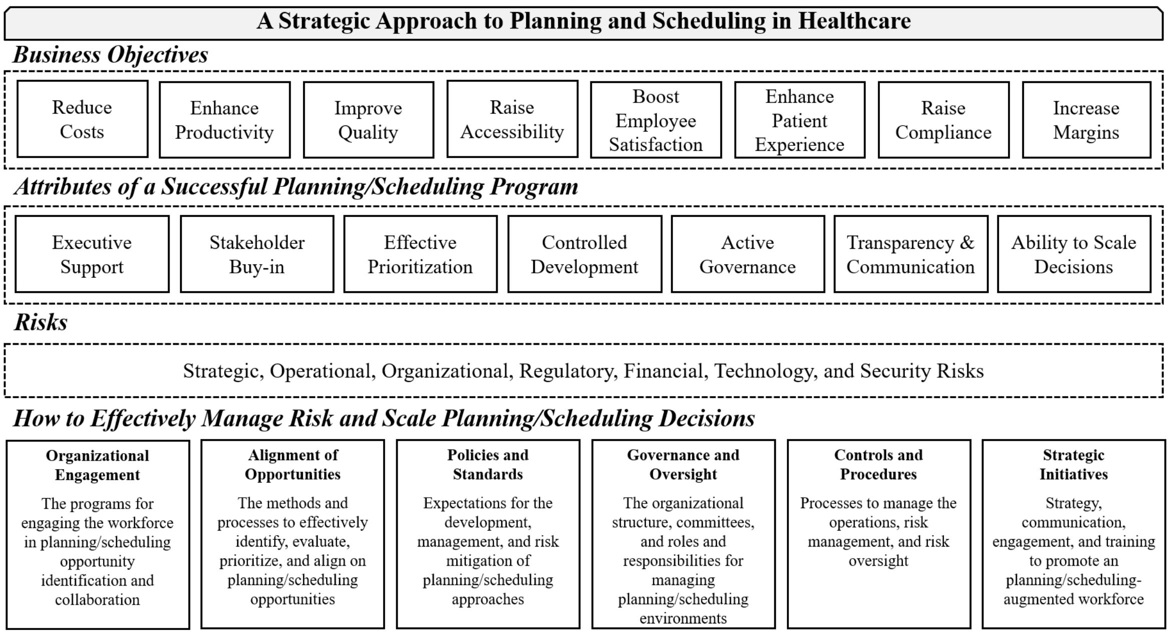
In this section, we further discuss emerging trends in healthcare and suggest potential research opportunities. For each research opportunity, we also include a list of related POM journal departments, other than *Healthcare Operations Management* and *Service Operations Management* departments, to promote multidisciplinary and multimethod studies.

4.1. Collaborative Ecosystem and Cooperative Competition Strategy

The pandemic has revealed deficiencies in healthcare organizations all across the world, including safety, equipment, data availability, and infrastructure. We learned during the pandemic that relying only on an organization's own supply lines and capacities could result in ad hoc alliances between providers, suppliers, and non-healthcare enterprises to bring resources and capacity to meet the crisis. A major takeaway is that successful businesses should look for opportunities to collaborate with partners that have complementary talents to resolve issues early in the strategic planning process. Cooperative competition, often known as *coopetition*, is a growing trend in healthcare. We provide three examples to illustrate when and why healthcare providers may adopt this coopetition strategy:

- **To focus on core specialty services for financial sustainability:** Basic primary care,

Figure 1 A Strategic Approach to Planning and Scheduling in Healthcare



simple diagnostic tests, and chronic disease management are now available at national drugstore chains such as CVS and Walmart. Existing health systems have frequently struggled to make money while providing these services. Identifying possibilities to work with retail companies to address this gap can help organizations streamline services, enhance access, and provide better patient care at a reduced cost by bridging the gap.

- **To increase market share while also improving community health:** New entrants can expand the total market for health services by acting as a force multiplier. About half of women in their forties and fifties, for example, do not obtain mammograms on a regular basis. The majority of women obtaining in-store mammograms would not require follow-up care if mammography services given by a prominent retailer were successful in motivating this population. Many, on the other hand, would require referrals for additional diagnostic tests and, maybe, therapy.[2] Establishing this two-way relationship with that new entrant, which includes sharing data and offering easy access to hospitals, might lead to a considerable increase in new referrals.
- **To relax the physical boundary of operations:** The talent pool is expanding, and new entrants are emerging that can offer services at a lower cost and frequently at a greater quality than is achievable for some businesses, thanks to the advent of virtual services and workforces, which we will discuss in more detail shortly. One example is a partnership between tele-ICU providers and small, rural hospitals to enhance access to highly specialized critical care for their patients.

[2] See https://trustees.aha.org/top-10-emerging-trends-health-care-2021-new-normal (Accessed date: Jan 5, 2022).

Prior literature discusses decision-making and consequences under limited resources, e.g., lack of inpatient beds causes denial of emergency patients' hospital admission (Kc and Terwiesch 2017), impacts of provider availability shocks on care channel diversion and delays (Bavafa et al. 2021), and physicians' decisions about ordering imaging tests (Dai et al. 2017). We encourage healthcare OM researchers to further investigate the capacity and scheduling implications of leveraging the cooperative competition strategy that helps increase downstream market capture and focuses on core specialty services while remaining highly connected to the patient.

Research Opportunity 1. *How and to what extent can the cooperative competition strategy and new planning/scheduling models be used to focus on financially sustainable services, expand the market, and leverage resources outside of physical boundaries?*

Related POM Departments: *New Business Models and Operations Innovation; Revenue Management and Market Analytics; Global Operations Strategy; Disaster Management; Supply Chain Management.*

4.2. Expanded Reach of Virtual Care

Virtual care solutions continue to emerge across the care continuum from telehealth visits to virtual hospital care and home-based care. Pre-pandemic, less than 2% of outpatient behavioral health and medical claims were for virtual visits. In 2021, they make up nearly 25%.[3] Experts expect to see more adoption and more innovations. Indeed, emerging virtual care organizations excel in several areas, such as urgent care, behavioral health therapy, dermatology, and wellness or primary care, including ordering lab tests and analyzing the results.[4] A good example is how telemedicine can be combined into therapy decisions. A study on continuous glucose monitoring and telemedicine visits for patients with type 1 and type 2 diabetes concluded that the use of telemedicine patient consults and remote monitoring of glucose and insulin data can significantly improve glycemic control.[5] Healthcare futurists even believe that by 2040 most care will be delivered at home, in outpatient settings, or virtually.[6]

Telehealth has the potential to save organizations money on healthcare while keeping employees healthy and productive. By reducing unnecessary office visits, telemedicine can save employees an expensive copay and lower claim expenses for the employer's group health plan. Surprisingly, however, Bavafa et al. (2018) discover that e-visits may lead to increased office visits, contrary to common expectations that they serve as a substitute. The authors demonstrate that systems at

[3] See https://www.cigna.com/static/www-cigna-com/docs/about-us/newsroom/innovation/cigna-covid-trends-factsheet-virtual-and-behavioral-health.pdf (Accessed date: Jan 9, 2022).

[4] See https://newsroom.cigna.com/4-biggest-health-care-trends-2022-impact-on-employers (Accessed date: Feb 2, 2022).

[5] https://medicalfuturist.com/emerging-trends-in-healthcare/ (Accessed date: Dec 10, 2021).

[6] See https://www2.deloitte.com/us/en/insights/industry/health-care/forces-of-change-health-care.html (Accessed date: Dec 10, 2021).

capacity may not reap the benefits of e-visits because increased visits by existing patients may reduce capacity for new patients without necessarily generating greater income. Meanwhile, Atasoy et al. (2021) find that health information exchanges (HIE) effectively reduce visits that are not preventative, while they increase adherence to preventative care visits, aiding in the shift from treatment-based care to preventative-based care. Naturally, future study should focus on better understanding the consequences of virtual care for various stakeholders, such as healthcare planning and scheduling managers, in diverse scenarios.

Research Opportunity 2. *How can we improve our use of virtual care to make care delivery safer, more productive, and efficient?*
Related POM Departments: *Management of Technology.*

4.3. Applications of Artificial Intelligence in Healthcare

While artificial intelligence (AI) is a growing area in operations management, many OM scholars have been exploring important research questions, using different methods such as queuing models, statistics, econometrics, and optimization, that are essential to developing and applying AI in healthcare. Example topics include using potential patient arrivals to effectively manage admissions into an ED (Xu and Chan 2016), measuring the higher risk of adverse patient outcomes caused by long delays (Chan et al. 2017), developing effective checkup plans to monitor patients following hospital discharge and reduce readmissions (Helm et al. 2016, Liu et al. 2018b), examining causes and impacts of doctors' deviation from their prescribed task sequence (Ibanez et al. 2018), structurally estimating ED patients' waiting cost (Ding et al. 2019), quantifying unwarranted clinical practice variation and exploring its impacts on operational performance (Youn et al. 2021), and calculating the impact of service facility layout on service workers' task organization using nurse location tracking data (Meng et al. 2021).

AI and automation are taking hold in healthcare at an accelerated rate as they have in other fields such as banking, media, and retail operations.[7] For example, AI-powered solutions based on the above literature can help with precisely scheduling and planning clinical staff rotation by taking in operational restrictions such the number of staff, availability, expertise, and specialized equipment needed. We list a few use cases below:

- **Quality and efficiency in radiology.** With solutions to reduce duplicate jobs, eliminate bias-based reading errors, discover data patterns in images to predict risk, and improve workflow procedures, AI is having a huge impact in radiology.
- **Real-time analytics to expedite care.** Real-time data is being used by organizations to drive the care process. For example, command center software solutions combine systems engineering, predictive analytics, and problem-solving to control patient flow in and out of the health system while ensuring clinical quality, safety, and the best possible patient experience.

[7] See https://trustees.aha.org/top-10-emerging-trends-health-care-2021-new-normal (Accessed date: Jan 10, 2022).

- **Productivity in nonclinical areas.** Healthcare operations that rely largely on repetitive processes, such as supply chain, revenue cycle, and customer support, are benefiting from AI-assisted automation.

Research Opportunity 3. *How can we leverage data and AI to boost productivity while also improving clinical outcomes?*

Related POM Departments: *Disruptive Technologies and Operations Management; POM-Information Systems Interface*

4.4. Patient Engagement and Personalized Care

The products we order today arrive the same day, and we can track them in real time from the time of order to delivery. It is not surprising that patients demand the same high level of efficiency and transparency from their healthcare providers. Meanwhile, some patients may have to wait weeks or months for an appointment, with no idea when their exam results will be available. To improve the convenience, speed, and transparency of care, healthcare organizations must assess their current barriers to consumer satisfaction and implement analytics and patient-centric technologies. For example, when a West Coast health system adopted precision scheduling practices to minimize wasted time between imaging exams, they were able to open up 5,000 extra exam slots each year, allowing patients to be seen sooner. [8] Other examples include understanding the effects of wait on the likelihood of no-shows (Osadchiy and Kc 2017), rescheduling on no-show behavior (Liu et al. 2019a), patient preferences and choice behavior on scheduling medical appointments (Liu et al. 2018a).

While patients desire the convenience and simplicity of digital interactions, personalized care remains the cornerstone of their devotion. An ideal healthcare experience, according to a 2020 survey of healthcare customers, requires a human touch, whether the interaction is virtual or in person. Patients say that clinicians must take time to listen, show they care, and communicate clearly.

Research Opportunity 4. *What types of clinicians or services should be considered in the planning stages to drive the level of experience that patients are demanding?*

Related POM Departments: *Behavioral Opeartions; POM-Marketing Interface*

4.5. Shifting to Value-Based Care

The Affordable Care Act was signed into law ten years ago in 2010, marking the beginning of the U.S. healthcare system's transformation from volume to value. The 10-year value-based care success report is mixed, as CMS and healthcare organizations are continually evolving and adapting policies and strategies to enhance outcomes and performance. While CMS's focus on quality and cost has resulted in significant reductions in readmissions, organizations still need to simplify and

[8] See https://trustees.aha.org/top-10-emerging-trends-health-care-2021-new-normal (Accessed date: Jan 10, 2022).

combine value-based programs to achieve more broad positive impacts.[9] To support the transition from volume to value, extensive studies have explored the impacts of various payment schemes and incentives on operational and financial performance. Examples include a performance-based contracting framework (Jiang et al. 2012), the role of a hospital's objective (i.e., non-profit vs. for-profit) in competitive hospital markets for elective care (Andritsos and Aflaki 2015), an analysis of subsidy schemes to reduce waiting times for public healthcare service (Qian et al. 2017), a health co-production model in which the patient's readmissions can be jointly controlled by both the hospital's and the patient's efforts under different payment schemes (Andritsos and Tang 2018), and several mechanisms (e.g., offering a value-based payment) that a payer such as Medicare could use to promote prehospital triage (Webb and Mills 2019).

Several other studies focus more on operational issues that are relevant in designing incentive structures that promote better patient experience and care outcomes. Examples include the impact of physician workload on hospital reimbursement (Powell et al. 2012), the ICU occupancy level and its relationship to patient discharege timing (Kc and Terwiesch 2012), the ICU patient admission process (Kim et al. 2015), the effect of patient arrival time on quality of care (Anderson et al. 2014), the influence of pooled versus dedicated queuing systems on patients' wait times and lengths of stay (Song et al. 2015), and the impact of HIE adoption on patients' lengths of stay (Ayer et al. 2019).

The healthcare industry is suspended between value-based and volume-based payment models, as rising costs push a shift toward value, but the current environment still favors volume. Hence, to thrive economically as value replaces volume, health systems must grasp the elements that drive and support both payment models and operations strategies.

Research Opportunity 5. *How and to what extent should the integration and interoperability of solutions, such electronic health records, imaging platforms, and clinical diagnostic decision support tools, be used in planning and scheduling so that hospitals can effectively meet value-based care goals while improving their own financial health?*

Related POM Departments: *Industry Studies & Public Policy; POM-Economics Interface; POM-Accounting Interface; POM-Finance Interface*

4.6. Working Toward Health Equity

We can support everyone's pursuit of optimal health by eradicating health inequities and removing unnecessary barriers to accessing the treatment they require. These inequalities, or social determinants of health, are systemic and can be based on a person's race, gender, sexual orientation, income, or place of residence.[10] For instance, there can be few or no hospitals or medical workers in rural areas or places with high rates of poverty. Health disparity can have serious repercussions. According

[9] See https://www.healthcatalyst.com/insights/value-based-purchasing-2020-10-year-progress-report/ (Accessed date: Jan 5, 2022).

[10] See https://newsroom.cigna.com/4-biggest-health-care-trends-2022-impact-on-employers (Accessed date: Feb 10, 2022).

to the Centers for Disease Control and Prevention, those without access to high-quality care are more likely to get ill or disabled and pass away at an earlier age. As reported in the Presidential COVID-19 Health Equity Task Force report in 2021, the epidemic has recently shed attention on health inequities in the U.S.[11] The report recommends a number of actions, many of which can be explored in strategic and operational levels of planning by healthcare OM researchers, including:

- Enforcing a data ecosystem that encourages equity-driven decision making
- Increasing accountability for health equity outcomes
- Investing in a representative healthcare workforce
- Investing in community-led solutions to address health inequity

Atasu et al. (2017), for example, examines how to manage the recovery of surplus (i.e., unused or donated) medical items from not-for-profit organizations to fulfill the needs of underprivileged healthcare facilities in developing countries.

Research Opportunity 6. *What are the roles and impacts of governments' and other agencies' policies and decision making structures to alleviate health disparities over broader healthcare networks?*

Related POM Departments: *Industry Studies & Public Policy; Disaster Management; Not-for-Profit Operations Management*

4.7. Aging Populations

The population is graying, and it is no surprise that older individuals utilize healthcare more frequently. All baby boomers will be beyond the age of 65 by 2030.[12] Naturally, this will put a strain on the healthcare system. According to the Centers for Disease Control and Prevention, while roughly 25% of Americans have numerous chronic illnesses, the rate rises to nearly 60% for those 65 and older.[13] The demand is growing, while the supply is shrinking. This is partly due to the fact that doctors are also becoming older. People aged 65 and up make up 15% of the workforce, while those aged 55 and 64 make up 27%. The Association of American Medical Colleges predicts that the U.S. will have a physician shortage of up to 121,900 by 2032.[14]

Hospitals may experience safety tipping points when workload variability buffers are depleted (Kuntz et al. 2015). One way to address the shortages and to support sustainable operations under aging populations might be teamwork. Increasingly, nurse practitioners and physician assistants are taking on many tasks traditionally performed by doctors, either as an "ice-breaker," seeing each

[11] See https://www.aamc.org/advocacy-policy/washington-highlights/presidential-covid-19-health-equity-task-force-releases-final-report (Accessed date: Feb 10, 2022).

[12] See https://www.wolterskluwer.com/en/expert-insights/3-emerging-healthcare-trends-that-will-shape-your-medical-career (Accessed date: Jan 22, 2022).

[13] See https://www.cdc.gov/mmwr/volumes/65/wr/mm6529a3.htm (Accessed date: Jan 22, 2022).

[14] See https://aamc-black.global.ssl.fastly.net/production/media/filer_public/31/13/3113ee5c-a038-4c16-89af-294a69826650/2019_update_-_the_complexities_of_physician_supply_and_demand_-_projections_from_2017-2032.pdf (Accessed date: Feb 10, 2022).

patient before the physician, or as a "standalone" provider, a substitute for the physician for the entirety of some patients' visits (White et al. 2017). Healthcare OM scholars have the expertise to help the industry properly adapt the business under such changing environments.

Research Opportunity 7. *How should healthcare systems prepare for an aging population and a growing disease burden?*

Related POM Departments: *Sustainable Operations; Non-for-Profit Operations Management*

5. Concluding Remarks

Planning and scheduling in healthcare is a complicated task, essentially because no two patients are alike, and all the stakeholders have to deal with numerous uncertain factors that are never possible to fully understand at the stages of planning and scheduling. By viewing such challenges as intriguing opportunities rather than obstacles, healthcare operations management researchers and practitioners have proposed and applied a variety of research methods over the past 30 years. We believe this journey will never end but will continue to evolve over the next 30 years.

The primary contribution of this paper is to synthesize the current state of the literature into several categories and to propose potentially important topics that could be explored in each subfield. Specifically, we categorized the capacity planning literature into a) hospital units and operating rooms, b) outpatient clinics, and c) broader healthcare networks. We further grouped the scheduling literature into a) appointment scheduling, b) surgery and workforce scheduling, c) recurring and integrated scheduling, with several subtopics that have attracted researchers' attention. Further, we demonstrate the seven emerging trends in healthcare (i.e., collaborative ecosystem and cooperative competition strategy, expanded reach of virtual care, applications of artificial intelligence in healthcare, patient engagement and personalized care, shifting to value-based care, working toward health equity, and aging populations) and identify potential research opportunities. In sum, research on planning and scheduling in healthcare is no longer simply an opportunity for individual researchers using a single method or focusing on a small application, rather it is a comprehensive area that should be pursued by multidisciplinary perspectives to deliver coherent solutions that incorporate the complex contemporary issues harmoniously.

Acknowledgments

The authors are grateful to the special issue editors Christopher S. Tang and Subodha Kumar for their valuable suggestions and constructive comments on this research.

References

Anderson, D., G. Gao, B. Golden. 2014. Life is all about timing: An examination of differences in treatment quality for trauma patients based on hospital arrival time. *Production and Operations Management* **23**(12) 2178–2190.

Andritsos, D. A., S. Aflaki. 2015. Competition and the operational performance of hospitals: The role of hospital objectives. *Production and Operations Management* **24**(11) 1812–1832.

Andritsos, D. A., C. S. Tang. 2018. Incentive programs for reducing readmissions when patient care is co-produced. *Production and Operations Management* **27**(6) 999–1020.

Armony, M., C. W. Chan, B. Zhu. 2018. Critical care capacity management: Understanding the role of a step down unit. *Production and Operations Management* **27**(5) 859–883.

Atasoy, H., E. M. Demirezen, P.-Y. Chen. 2021. Impacts of patient characteristics and care fragmentation on the value of hies. *Production and Operations Management* **30**(2) 563–583.

Atasu, A., B. Toktay, W. M. Yeo, C. Zhang. 2017. Effective medical surplus recovery. *Production and Operations Management* **26**(6) 1142–1162.

Ayer, T., M. U. S. Ayvaci, Z. Karaca, J. Vlachy. 2019. The impact of health information exchanges on emergency department length of stay. *Production and Operations Management* **28**(3) 740–758.

Barz, C., K. Rajaram. 2015. Elective patient admission and scheduling under multiple resource constraints. *Production and Operations Management* **24**(12) 1907–1930.

Bavafa, H., A. Canamucio, S. C. Marcus, C. Terwiesch, R. M. Werner. 2021. Capacity rationing in primary care: Provider availability shocks and channel diversion. *Management Science,* Forthcoming.

Bavafa, H., L. M. Hitt, C. Terwiesch. 2018. The impact of e-visits on visit frequencies and patient health: Evidence from primary care. *Management Science* **64**(12) 5461–5480.

Bavafa, H., S. Savin, C. Terwiesch. 2019. Managing patient panels with non-physician providers. *Production and Operations Management* **28**(6) 1577–1593.

Best, T. J., B. Sandıkçı, D. D. Eisenstein, D. O. Meltzer. 2015. Managing hospital inpatient bed capacity through partitioning care into focused wings. *Manufacturing & Service Operations Management* **17**(2) 157–176.

Bidhandi, Hadi Mohammadi, Jonathan Patrick, Pedram Noghani, Peyman Varshoei. 2019. Capacity planning for a network of community health services. *European Journal of Operational Research* **275**(1) 266–279.

Bretthauer, K. M., H. S. Heese, H. Pun, E. Coe. 2011. Blocking in healthcare operations: A new heuristic and an application. *Production and Operations Management* **20**(3) 375–391.

Cardoen, B., E. Demeulemeester, J. Beliën. 2010. Operating room planning and scheduling: A literature review. *European Journal of Operational Research* **201**(3) 921–932.

Carew, S., M. Nagarajan, S. Shechter, J. Arneja, E. Skarsgard. 2021. Dynamic capacity allocation for elective surgeries: Reducing urgency-weighted wait times. *Manufacturing & Service Operations Management* **23**(2) 407–424.

Cayirli, T., E. Veral. 2003. Outpatient scheduling in health care: a review of literature. *Production and Operations Management* **12**(4) 519–549.

Cayirli, T., E. Veral, H. Rosen. 2008. Assessment of patient classification in appointment system design. *Production and Operations Management* **17**(3) 338–353.

Cayirli, T., K. K. Yang, S. A. Quek. 2012. A universal appointment rule in the presence of no-shows and walk-ins. *Production and Operations Management* **21**(4) 682–697.

Chan, C. W., V. F. Farias, G. J. Escobar. 2017. The impact of delays on service times in the intensive care unit. *Management Science* **63**(7) 2049–2072.

Chan, C. W., M. Huang, V. Sarhangian. 2021. Dynamic server assignment in multiclass queues with shifts, with applications to nurse staffing in emergency departments. *Operations Research,* Forthcoming.

Chen, R. R., L. W. Robinson. 2014. Sequencing and scheduling appointments with potential call-in patients. *Production and Operations Management* **23**(9) 1522–1538.

Cherkesly, M., M.-È. Rancourt, K. R. Smilowitz. 2019. Community healthcare network in underserved areas: design, mathematical models, and analysis. *Production and Operations Management* **28**(7) 1716–1734.

Childers, C. P., M. Maggard-Gibbons. 2018. Understanding costs of care in the operating room. *JAMA Surgery* **153**(4) e176233–e176233.

Chow, V. S., M. L. Puterman, N. Salehirad, W. Huang, D. Atkins. 2011. Reducing surgical ward congestion through improved surgical scheduling and uncapacitated simulation. *Production and Operations Management* **20**(3) 418–430.

Cire, A. A., A. Diamant, T. Yunes, A. Carrasco. 2019. A network-based formulation for scheduling clinical rotations. *Production and Operations Management* **28**(5) 1186–1205.

Dai, T., M. Akan, S. Tayur. 2017. Imaging room and beyond: The underlying economics behind physicians' test-ordering behavior in outpatient services. *Manufacturing & Service Operations Management* **19**(1) 99–113.

Day, R., R. Garfinkel, S. Thompson. 2012. Integrated block sharing: a win–win strategy for hospitals and surgeons. *Manufacturing & Service Operations Management* **14**(4) 567–583.

Deglise-Hawkinson, J., J. E. Helm, T. Huschka, D. L. Kaufman, M. P. Van Oyen. 2018. A capacity allocation planning model for integrated care and access management. *Production and Operations Management* **27**(12) 2270–2290.

Denton, B. T., A. J. Miller, H. J. Balasubramanian, T. R. Huschka. 2010. Optimal allocation of surgery blocks to operating rooms under uncertainty. *Operations Research* **58**(4-part-1) 802–816.

Diamant, A., J. Milner, F. Quereshy. 2018. Dynamic patient scheduling for multi-appointment health care programs. *Production and Operations Management* **27**(1) 58–79.

Ding, Y., E. Park, M. Nagarajan, E. Grafstein. 2019. Patient prioritization in emergency department triage systems: An empirical study of the canadian triage and acuity scale (ctas). *Manufacturing & Service Operations Management* **21**(4) 723–741.

Dobson, G., S. Hasija, E. J. Pinker. 2011. Reserving capacity for urgent patients in primary care. *Production and Operations Management* **20**(3) 456–473.

Easton, F. F., D. F. Rossin, W. S. Borders. 1992. Analysis of alternative scheduling policies for hospital nurses. *Production and Operations Management* **1**(2) 159–174.

Feldman, J., N. Liu, H. Topaloglu, S. Ziya. 2014. Appointment scheduling under patient preference and no-show behavior. *Operations Research* **62**(4) 794–811.

Flinkman, M., H. Leino-Kilpi, S. Salanterä. 2010. Nurses' intention to leave the profession: integrative review. *Journal of Advanced Nursing* **66**(7) 1422–1434.

Forbes, S. H. Jain. 2019. Missed appointments, missed opportunities: tackling the patient no-show problem URL https://www.forbes.com/sites/sachinjain/2019/10/06/missed-appointments-missed-opportunities-tackling-the-patient-no-show-problem/?sh=25dc0d9e573b. Accessed: 03-10-2022.

Freeman, N. K., S. H. Melouk, J. Mittenthal. 2016. A scenario-based approach for operating theater scheduling under uncertainty. *Manufacturing & Service Operations Management* **18**(2) 245–261.

Gökalp, E., N. Gülpınar, X. V. Doan. 2020. Capacity planning for networks of stem-cell donation centers under uncertainty. *Production and Operations Management* **29**(2) 281–297.

Grant, B., I. Gurvich, R. K. Mutharasan, J. A. Van Mieghem. 2022. Optimal dynamic appointment scheduling of base and surge capacity. *Manufacturing & Service Operations Management* **24**(1) 59–76.

Green, L. V., S. Savin, B. Wang. 2006. Managing patient service in a diagnostic medical facility. *Operations Research* **54**(1) 11–25.

Gul, S., B. T. Denton, J. W. Fowler, T. Huschka. 2011. Bi-criteria scheduling of surgical services for an outpatient procedure center. *Production and Operations Management* **20**(3) 406–417.

Gupta, D. 2007. Surgical suites' operations management. *Production and Operations Management* **16**(6) 689–700.

Gupta, D., B. Denton. 2008. Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions* **40**(9) 800–819.

Gupta, D., S. J. Potthoff, et al. 2016. Matching supply and demand for hospital services. *Foundations and Trends® in Technology, Information and Operations Management* **8**(3–4) 131–274.

Gupta, D., L. Wang. 2008. Revenue management for a primary-care clinic in the presence of patient choice. *Operations Research* **56**(3) 576–592.

Gupta, S., M. K. Starr, R. Z. Farahani, N. Asgari. 2021. Pandemics/epidemics: Challenges and opportunities for operations management research. *Manufacturing & Service Operations Management,* Forthcoming.

Halter, M., O. Boiko, F. Pelone, C. Beighton, R. Harris, J. Gale, S. Gourlay, V. Drennan. 2017. The determinants and consequences of adult nursing staff turnover: a systematic review of systematic reviews. *BMC Health Services Research* **17**(1) 1–20.

Helm, J. E., S. AhmadBeygi, M. P. Van Oyen. 2011. Design and analysis of hospital admission control for operational effectiveness. *Production and Operations Management* **20**(3) 359–374.

Helm, J. E., A. Alaeddini, J. M. Stauffer, K. M. Bretthauer, T. A. Skolarus. 2016. Reducing hospital readmissions by integrating empirical prediction with resource optimization. *Production and Operations Management* **25**(2) 233–257.

Helm, J. E., M. P. Van Oyen. 2014. Design and optimization methods for elective hospital admissions. *Operations Research* **62**(6) 1265–1282.

Ibanez, M. R., J. R. Clark, R. S. Huckman, B. R. Staats. 2018. Discretionary task ordering: Queue management in radiological services. *Management Science* **64**(9) 4389–4407.

Izady, N. 2015. Appointment capacity planning in specialty clinics: A queueing approach. *Operations Research* **63**(4) 916–930.

Jiang, H., Z. Pang, S. Savin. 2012. Performance-based contracts for outpatient medical services. *Manufacturing & Service Operations Management* **14**(4) 654–669.

Jung, K. S., M. Pinedo, C. Sriskandarajah, V. Tiwari. 2019. Scheduling elective surgeries with emergency patients at shared operating rooms. *Production and Operations Management* **28**(6) 1407–1430.

Kc, D. S., C. Terwiesch. 2012. An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing & Service Operations Management* **14**(1) 50–65.

Kc, D. S., C. Terwiesch. 2017. Benefits of surgical smoothing and spare capacity: an econometric analysis of patient flow. *Production and Operations Management* **26**(9) 1663–1684.

Keskinocak, P., N. Savva. 2020. A review of the healthcare-management (modeling) literature published in manufacturing & service operations management. *Manufacturing & Service Operations Management* **22**(1) 59–72.

Kim, S.-H., C. W. Chan, M. Olivares, G. Escobar. 2015. Icu admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Science* **61**(1) 19–38.

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

Kong, Q., S. Li, N. Liu, C.-P. Teo, Z. Yan. 2020. Appointment scheduling under time-dependent patient no-show behavior. *Management Science* **66**(8) 3480–3500.

Kucukyazici, B., V. Verter, N. E. Mayo. 2011. An analytical framework for designing community-based care for chronic diseases. *Production and Operations Management* **20**(3) 474–488.

Kuntz, L., R. Mennicken, S. Scholtes. 2015. Stress on the ward: Evidence of safety tipping points in hospitals. *Management Science* **61**(4) 754–771.

LaGanga, L. R., S. R. Lawrence. 2012. Appointment overbooking in health care clinics to improve patient service and clinic performance. *Production and Operations Management* **21**(5) 874–888.

Lee, D. K. K., S. A. Zenios. 2009. Optimal capacity overbooking for the regular treatment of chronic conditions. *Operations Research* **57**(4) 852–865.

Lee, S. J., G. R. Heim, C. Sriskandarajah, Y. Zhu. 2018. Outpatient appointment block scheduling under patient heterogeneity and patient no-shows. *Production and Operations Management* **27**(1) 28–48.

Lemay, B., A. Cohn, M. Epelman, S. Gorga. 2017. New methods for resolving conflicting requests with examples from medical residency scheduling. *Production and Operations Management* **26**(9) 1778–1793.

Li, Y., S. Y. Tang, J. Johnson, D. A. Lubarsky. 2019. Individualized no-show predictions: Effect on clinic overbooking and appointment reminders. *Production and Operations Management* **28**(8) 2068–2086.

Liu, J., J. Xie, K. K. Yang, Z. Zheng. 2019a. Effects of rescheduling on patient no-show behavior in outpatient clinics. *Manufacturing & Service Operations Management* **21**(4) 780–797.

Liu, N. 2016. Optimal choice for appointment scheduling window under patient no-show behavior. *Production and Operations Management* **25**(1) 128–142.

Liu, N., S. R. Finkelstein, M. E. Kruk, D. Rosenthal. 2018a. When waiting to see a doctor is less irritating: Understanding patient preferences and choice behavior in appointment scheduling. *Management Science* **64**(5) 1975–1996.

Liu, N., V.-A. Truong, X. Wang, B. R. Anderson. 2019b. Integrated scheduling and capacity planning with considerations for patients' length-of-stays. *Production and Operations Management* **28**(7) 1735–1756.

Liu, N., S. Ziya. 2014. Panel size and overbooking decisions for appointment-based services under patient no-shows. *Production and Operations Management* **23**(12) 2209–2223.

Liu, X., M. Hu, J. E. Helm, M. S. Lavieri, T. A. Skolarus. 2018b. Missed opportunities in preventing hospital readmissions: Redesigning post-discharge checkup policies. *Production and Operations Management* **27**(12) 2226–2250.

Lovejoy, W. S., Y. Li. 2002. Hospital operating room capacity expansion. *Management Science* **48**(11) 1369–1387.

Mandelbaum, A., P. Momčilović, N. Trichakis, S. Kadish, R. Leib, C. A. Bunnell. 2020. Data-driven appointment-scheduling under uncertainty: The case of an infusion unit in a cancer center. *Management Science* **66**(1) 243–270.

Martin, A. B., M. Hartman, B. Washington, A. Catlin, National Health Expenditure Accounts Team, et al. 2019. National health care spending in 2017: growth slows to post–great recession rates; share of gdp stabilizes. *Health Affairs* 10–1377.

May, J. H., W. E. Spangler, D. P. Strum, L. G. Vargas. 2011. The surgical scheduling problem: Current research and future opportunities. *Production and Operations Management* **20**(3) 392–405.

McCoy, J. H., E. M. Johnson. 2014. Clinic capacity management: Planning treatment programs that incorporate adherence. *Production and Operations Management* **23**(1) 1–18.

Meng, F., J. Qi, M. Zhang, J. Ang, S. Chu, M. Sim. 2015. A robust optimization model for managing elective admission in a public hospital. *Operations Research* **63**(6) 1452–1467.

Meng, Lesley, Robert J Batt, Christian Terwiesch. 2021. The impact of facility layout on service worker behavior: An empirical study of nurses in the emergency department. *Manufacturing & Service Operations Management* .

Mills, A. F., J. E. Helm, Y. Wang. 2021. Surge capacity deployment in hospitals: effectiveness of response and mitigation strategies. *Manufacturing & Service Operations Management* **23**(2) 367–387.

Naderi, B., V. Roshanaei, M. A. Begen, D. M. Aleman, D. R. Urbach. 2021. Increased surgical capacity without additional resources: Generalized operating room planning and scheduling. *Production and Operations Management,* Forthcoming.

Osadchiy, N., D. Kc. 2017. Are patients patient? the role of time to appointment in patient flow. *Production and Operations Management* **26**(3) 469–490.

Patrick, J. 2011. Access to long-term care: The true cause of hospital congestion? *Production and Operations Management* **20**(3) 347–358.

Patrick, J., M. L. Puterman, M. Queyranne. 2008. Dynamic multipriority patient scheduling for a diagnostic resource. *Operations Research* **56**(6) 1507–1525.

Pinedo, M. 2005. *Planning and Scheduling in Manufacturing and Services*. Springer.

Powell, A., S. Savin, N. Savva. 2012. Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing & Service Operations Management* **14**(4) 512–528.

Price, C., B. Golden, M. Harrington, R. Konewko, E. Wasil, W. Herring. 2011. Reducing boarding in a post-anesthesia care unit. *Production and Operations Management* **20**(3) 431–441.

Qian, Q., P. Guo, R. Lindsey. 2017. Comparison of subsidy schemes for reducing waiting times in healthcare systems. *Production and Operations Management* **26**(11) 2033–2049.

Rath, S., K. Rajaram, A. Mahajan. 2017. Integrated anesthesiologist and room scheduling for surgeries: Methodology and application. *Operations Research* **65**(6) 1460–1478.

Roth, A. V., R. Van Dierdonck. 1995. Hospital resource planning: concepts, feasibility, and framework. *Production and Operations Management* **4**(1) 2–29.

Salzarulo, P. A., K. M. Bretthauer, M. J. Côté, K. L. Schultz. 2011. The impact of variability and patient information on health care system performance. *Production and Operations Management* **20**(6) 848–859.

Salzarulo, P. A., S. Mahar, S. Modi. 2016. Beyond patient classification: Using individual patient characteristics in appointment scheduling. *Production and Operations Management* **25**(6) 1056–1072.

Samorani, M., S. Ganguly. 2016. Optimal sequencing of unpunctual patients in high-service-level clinics. *Production and Operations Management* **25**(2) 330–346.

Smith-Daniels, V. L., S. B. Schweikhart, D. E. Smith-Daniels. 1988. Capacity management in health care services: Review and future research directions. *Decision Sciences* **19**(4) 889–919.

Song, H., A. L. Tucker, K. L. Murrell. 2015. The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Science* **61**(12) 3032–3053.

Thompson, S., M. Nunez, R. Garfinkel, M. D. Dean. 2009. Or practice—efficient short-term allocation and reallocation of patients to floors of a hospital during demand surges. *Operations Research* **57**(2) 261–273.

Truong, V.-A. 2015. Optimal advance scheduling. *Management Science* **61**(7) 1584–1597.

Van Mieghem, J. A. 2003. Commissioned paper: Capacity management, investment, and hedging: Review and recent developments. *Manufacturing & Service Operations Management* **5**(4) 269–302.

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

Wang, D., K. Muthuraman, D. Morrice. 2019. Coordinated patient appointment scheduling for a multistation healthcare network. *Operations Research* **67**(3) 599–618.

Wang, W.-Y., D. Gupta. 2011. Adaptive appointment systems with patient preferences. *Manufacturing & Service Operations Management* **13**(3) 373–389.

Wang, W.-Y., D. Gupta. 2014. Nurse absenteeism and staffing strategies for hospital inpatient units. *Manufacturing & Service Operations Management* **16**(3) 439–454.

Webb, E. M., A. F. Mills. 2019. Incentive–compatible prehospital triage in emergency medical services. *Production and Operations Management* **28**(9) 2221–2241.

White, D. L., C. M. Froehle, K. J. Klassen. 2011. The effect of integrated scheduling and capacity policies on clinical efficiency. *Production and Operations Management* **20**(3) 442–455.

White, D. L., E. Torabi, C. M. Froehle. 2017. Ice-breaker vs. standalone: Comparing alternative workflow modes of mid-level care providers. *Production and Operations Management* **26**(11) 2089–2106.

Xu, K., C. W. Chan. 2016. Using future information to reduce waiting times in the emergency department via diversion. *Manufacturing & Service Operations Management* **18**(3) 314–331.

Youn, S., H. N. Geismar, C. Sriskandarajah, T. Vikram. 2022. Adaptive capacity planning for ambulatory surgery centers. *Manufacturing & Service Operations Management,* Forthcoming.

Youn, Seokjun, Gregory R Heim, Subodha Kumar, Chelliah Sriskandarajah. 2021. Examining the impacts of clinical practice variation on operational performance. *Production and Operations Management* **30**(4) 839–863.

Yu, S., V. G. Kulkarni, V. Deshpande. 2020. Appointment scheduling for a health care facility with series patients. *Production and Operations Management* **29**(2) 388–409.

Zacharias, C., M. Pinedo. 2014. Appointment scheduling with no-shows and overbooking. *Production and Operations Management* **23**(5) 788–801.

Zaerpour, F., M. Bijvank, H. Ouyang, Z. Sun. 2022. Scheduling of physicians with time-varying productivity levels in emergency departments. *Production and Operations Management* **31**(2) 645–667.

Zhou, Y., M. Parlar, V. Verter, S. Fraser. 2021. Surgical scheduling with constrained patient waiting times. *Production and Operations Management* **30**(9) 3253–3271.

Zhu, S., W. Fan, S. Yang, J. Pei, P. M. Pardalos. 2019. Operating room planning and surgical case scheduling: a review of literature. *Journal of Combinatorial Optimization* **37**(3) 757–805.