Published in final edited form as:

IIE Trans. 2020 ; 52(8): 832–849. doi:10.1080/24725854.2019.1675202.

Access Planning and Resource Coordination for Clinical Research Operations

Jivan Deglise-Hawkinson¹, David L. Kaufman², Blake Roessler^{3,4}, Mark P. Van Oyen¹

¹Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI

²Management Studies, University of Michigan–Dearborn, Dearborn, MI

³Department of Internal Medicine, University of Michigan, Ann Arbor, MI

⁴Michigan Clinical Research Unit (MCRU), University of Michigan, Ann Arbor, MI

Abstract

This research creates an operations engineering and management methodology to optimize a complex operational planning and coordination challenge faced by sites that perform clinical research trials. The time-sensitive and resource-specific treatment sequences for each of the many trial protocols conducted at a site make it very difficult to capture the dynamics of this unusually complex system. Existing approaches for site planning and participant scheduling exhibit both excessively long and highly variable Time to First Available Visit (TFAV) waiting times and high staff overtime costs. We have created a new method, termed CAPacity Planning Tool And INformatics (CAPTAIN) that provides decision support to identify the most valuable set of research trials to conduct within available resources and a plan for how to book their participants. Constraints include (i) the staff overtime costs, and/or (ii) the TFAV by trial. To estimate the site's metrics via a Mixed Integer Program, CAPTAIN combines a participant trajectory forecasting with an efficient visit booking reservation plan to allocate the date for the first visit of every participant's treatment sequence. It also plans a daily nursing staff schedule that is optimized together with the booking reservation plan to optimize each nurse's shift assignments in consideration of participants' requirements/needs.

Keywords

Health Care Operations; Patient Flow Modeling; Mixed Integer Programming; Production Planning and Scheduling; Stochastic Models

1 Introduction

Clinical research trials are an essential part of discovering and testing new medical devices, procedures, and drugs. This paper develops a methodology for appointment visit planning to improve clinical research trial delivery in dedicated performance sites, as a gateway to better scientific knowledge, medical discoveries, and effective new treatments for many diseases.

As described by Rubin et al. (2007), “These trials, research studies carried out with human volunteers to answer specific questions concerning the effectiveness of a drug, device, treatment or diagnostic method, are designed to advance scientific knowledge and promote discoveries to treat and cure illness and disease, and increase longevity and the quality of life for countless people”. While clinical trials are conducted in various settings, academic, medical, and industrial institutions frequently choose to construct clinical research performance sites to provide a shared infrastructure on which to perform them. We will use the term Clinical Research Unit (CRU) to refer to such an organization/site, including our partnering site, the Michigan Clinical Research Unit (MCRU). Operationally, a CRU may benefit from economies of scale and scope by sharing resources across many trials. On the other hand, some clinical trials have a complex operational structure that makes it difficult to plan and schedule them at a shared CRU site. “The conduct of a clinical trial is a complex integration of many activities requiring the coordination of a large number of individuals each with specific expertise,” states Heine et al. (2003). Unfortunately, as observed by Farrell et al. (2010), “Many clinical trials fail to deliver because of the lack of a structured, practical, businesslike approach to trial management.” (See Johansen et al. (1991) for further insight.) Until now though, rigorous research as well as the software vendor community have left the topic of CRU operations almost untouched. Our research addresses this problem by developing a strategic decision support system called CAPacity Planning Tool And INformatics (CAPTAIN) for better resource planning and coordination, faster access to treatment, efficient booking, and trial portfolio selection decision support in CRUs.

1.1 The complexity of CRU operations

To gain an understanding of a typical CRU, it is useful to consider the perspectives of some key stakeholders. First, participants, who may or may not be accurately called patients, can have many motives to participate in a trial, including the potential to receive a new treatment that is better than existing ones (although this would be speculative). For a participant who is screened and found to be appropriate to proceed in a trial, there will be a protocol specifying the details of how the trial must be conducted. “Clinical trials are detailed plans for medical treatments, for instance a clinical trial protocol can describe at which point of time, in which quantum, and how medications or therapies have to be executed,” states Heine et al. (2003).

Second, individual research investigators and their trial coordinators can be viewed as customers of the CRU who own the trials and work closely with CRU staff to administer the trials. Trial coordinators will follow a participant recruitment plan (based on judgment and past recruitment efforts for similar trials) from the initiation of the trial until either a designated number of participants has been reached or, in some cases, until a trial end date has been reached. Hence, participant enrollments (the times at which participants are screened and join the trial) are stochastic over time.

Third, operational planners for the CRU control the portfolio of clinical trials that they accept and try to balance the competing needs for service of the trials/protocols. CRU operational planners must choose their portfolio of clinical trials under complex resource constraints, without knowing with certainty what the later realized enrollments will be. The coordinators of the chosen trials are then responsible for recruiting the participants. Key

managerial questions faced by the operational planners include the following: • Given a set of candidate trials, what is the optimal portfolio composition? • If some new trials are accepted, what will be the impact on the available resources, given the incumbent clinical trials already underway? • Is it even feasible to accept some proposed trial? • For a portfolio of clinical trials, how can resources best be used? CAPTAIN is designed to both (1) aid the CRU operational planners in their portfolio selection and (2) ensure that the resources are utilized efficiently.

The main operational features of a single clinical trial include: (i) participant enrollments are stochastic over time, (ii) each participant enrolling in a trial has a multi-visit treatment path (often between 1 and 10 visits) with nearly deterministic service times that are specified by the trial's protocol, (iii) the visits of the treatment path are time-sensitive and can range from a few days to a year from enrollment within specific time windows (for example, visit 3 of a trial has to happen between 29 and 31 days after the first visit), and (iv) each trial visit requires a number of specific resources – chairs, beds, specific rooms/equipment, and specific skills of the nursing staff. We note that the determinism of service times in feature (ii) is realistic, and not just a convenient modelling assumption, because treatments are delivered according to precise protocols.

There are many facets to the protocols, physical resources, and human resources. For example, MCRU has to take into account nine different nursing skills: the ability to take vital signs, single venipuncture, IV catheter placement, simple intravenous infusions/injections, complex IV infusions and blood draws, frequent blood draws out of IV catheters, procedure monitoring, and subcutaneous injections and oral medication monitoring. While all nurses can take vital signs, only 80% can perform IV catheter placements, and a different 80% of the nurses can perform simple intravenous infusions. Fewer yet, 70%, can perform complex IV infusions and blood draws; etc. As for physical resources, some protocols can be done in any room with a chair or bed. Others specifically require procedure rooms with beds. Some require specialized equipment like DEXA scanners; etc. The protocols are highly variable in terms of the number of visits and the amount of time in between. The key dimensions of the problem are presented in Table 1 (Sec. 4.1). The many facets make it a challenge to describe and understand the model and its behavior. Therefore, we present a much simpler, albeit stylized, example in Appendix A that demonstrates portfolio selection tradeoffs and sensitivities.

One difficulty is that the scheduling of all of a participant's visits should be done on the enrollment date to ensure that the capacity is available to avoid protocol violations. That is, at that time there should be a check that the trial protocol's time and resource requirements will be met for all future visits. Each research trial must have a sufficient number of participants to achieve statistical validity. Either missing a visit or experiencing excessive delay may invalidate that participant, so it is important to ensure that the CRU's care providers and physical resources are available to service the scheduled participants. While uncommon, there has been some effort to develop an information system to identify the key human and physical resources required to fulfill the visits that are on the schedule, and to check for a resource conflict before allowing a centralized CRU scheduler to make an appointment. However, the complex resource and time coordination needs of the many trials

of a CRU cannot be managed well under current planning and scheduling methods. CAPTAIN was developed to meet an unfilled need to (i) forecast and control key CRU performance metrics (e.g., avoiding high overtime costs), (ii) perform the cost/benefit tradeoffs of accepting/starting a particular new trial, and (iii) generate a plan for how to keep participant access/wait times to an acceptable limit.

A key metric is the “Time to First Available Visit” (TFAV). We define TFAV as the time from a participant’s enrollment date to their first visit. In clinical research practice, the longer the TFAV, the greater the fraction of participants who choose not to start the trial, thus increasing overall trial duration due to a lower enrollment rate. High TFAV is also a problem for participants that need to start the trial as soon as possible due to their clinical condition (e.g., advanced cancer). Furthermore, as stated by Roland and Litka (2007), “each day it takes to bring [drug] products to the market is revenue lost: therefore, planning is critical to the entire drug development process, including clinical trial operations.” Trial duration is a major problem in current practice.

One of the main contributions of our CAPTAIN methodology is to introduce a new mechanism for resource booking, which we call the “Booking Reservation Plan” (BRPlan). The BRPlan reserves resource capacity for first visits of a treatment sequence, segmented by participant type. This is in contrast to the current practice booking methodology, which we call “First Available Slot” (FAS) dynamic scheduling, that schedules first visits into the first time slot that feasibly allows all of the future visits of the trial to be scheduled within specified time windows, irrespective of participant type. In order to understand the potential benefits of the BRPlan capacity reservation, it helps to first gain insights into the current FAS booking methodology, which we describe in detail next.

1.2 High level insights into current CRU functioning

When a new trial is proposed to the CRU, the managerial staff has to make the decision of whether or not the scientific/financial value they will gain from accepting it will outweigh the resources it requires (human as well as room), or whether or not there are even enough resources to accept it. Perhaps more importantly, there is currently no way to accurately determine the negative impact of a new trial imposed on existing trials (i.e., externalities) due to the increased congestion in the system.

CRU performance is dominated by the methods used to schedule the first visit of the treatment sequence required by the trial protocol, since subsequent visit times are dictated in the protocol from the time of the first visit. Failure to meet the required time window and resource requirements for any visit after the first results in a protocol violation that devalues research data and wastes resources. Furthermore, excessive violations will compromise the scientific integrity of the clinical trial if the target enrollment cannot be met. If all visits are properly booked at the time the participant enrolls, double booked appointments and congestion, which may result in protocol violations or overutilization, are avoided. As a rule, the managerial staff of a CRU would rather delay the start date of a participant’s treatment sequence (or perform some visits in overtime) than risk a protocol violation. In current practice and in our planning model, the violation rate metric is approximately 0.

Fig. 1 depicts an example of the current practice FAS booking methodology. We suppose that on day t a trial coordinator contacts the CRU to enroll a participant in a given trial. Suppose that this trial requires 3 visits: the second visit has to be scheduled between 10 and 12 days after the first visit, and the third visit has to be scheduled exactly 16 days after the first visit. The trial coordinator will communicate with the CRU to schedule this participant as soon as possible for their first visit. But this can only be done if there will be a nurse available with the required skills (this includes not only the nurses' nominal daily capacities but also the extra capacity that the CRU is willing to spend on overutilization) for all visits of the trial. The same goes for room/equipment availability. In Fig. 1, there are no nurses with the right skills available to schedule this participant for their first visit on day $t + 1$. On day $t + 2$, there is a nurse available with the required skills for visit 1 of such a trial. However, no nurse with the right skills for visit 2 is available from day $t + 11$ to day $t + 13$ (10 to 12 days after the first visit). There is no choice but to try to reschedule the participant's first visit to day $t + 3$. The same procedure is repeated (as seen in Fig. 1) until this participant is finally able to be scheduled on day $t + 5$ for a TFAV of 5 days.

Conceptually, it adds insight to consider the idea that a relatively more “complex” protocol contributes more to the TFAV than does a simpler one. Loosely speaking, they cause more congestion to the system, and they also suffer more from congestion, experiencing relatively poorer access. Our intuition suggests that the complexity of a protocol increases (*ceteris paribus*) with (i) a larger number of visits in the trial, (ii) smaller visit time windows, (iii) shorter times between visits (i.e., the rate of resource consumption increases), and (iv) larger workloads placed on the CRU. Such “higher density” protocols tend to have a long and more variable TFAV. This effect has been observed in many healthcare settings and is elucidated as follows: if poor capacity planning either utilizes too little capacity (i.e., is inefficient) or permits too many protocols to be conducted, then the system will experience significant congestion, thereby delaying the starting times of participants in every trial (increasing TFAV).

The more congested the system, the more difficult it is to schedule multi-visit participants, especially for high density protocols. We use simulation to demonstrate this behavior under FAS. The left of Fig. 2 displays the TFAV population average (averaged over the population of participants, subject to simulation error) and population standard deviation of 16 trials with protocol characteristics populated by one problem instance of a randomized test suite described in Sec. 4 (Table 1). The estimates are good enough to avoid the need for error bars. These 16 trials form a portfolio and are analyzed jointly. In the figure these are labeled as “active” trials. We will use the term active to refer to incumbent trials that are already active and cannot be taken out of a portfolio. Later, in Sec. 4.2, we will consider the optimization problem of selecting new trials from a set of “candidate” trials to potentially be added to these active trials to form a new, larger portfolio.

For congested systems, we find highly variable average TFAVs (average TFAV for a simple trial vs. a dense trial) as well as high variability in TFAV for participants within the same trial (especially in high density protocols). A main take-away is that, under FAS, there can be very different access delays for different trials, and there is a lot of variability around the averages. The right of Fig. 2 shows the effects on the population average TFAV when the

average workload per visit of trial number 3 increases; i.e., the trial's density increases (the number of visits is fixed). Trial 3 was selected here because its workload rate of 3.0 hours per visit happens to be close to the workload rate average across the 16 trials, and at that rate its population average TFAV is nearly the same as the average of the other trials – right down the middle, in both workload rate and average TFAV (see the intersection of the lines on the right of Fig. 2). We used the same problem instance as in the left of Fig. 2 while varying the workload per visit of trial 3. Trial 3's enrollment rate constitutes roughly 8% of the total portfolio enrollment rate in the left graph, and the point with a workload rate of 3.0 hours per visit in the right graph corresponds to its value in the base case. First, as expected, the population average TFAV of trial 3 increases nonlinearly in the average workload per visit. Also, there is not only a non-linear increase in the population average TFAV over all trials in the portfolio excluding trial 3, but the standard deviation bars on the means show increasing variability around this TFAV. This shows how the density of one trial not only affects the TFAV for its own participants, but greatly affects the TFAV of participants of all other trials (since the portfolio is conducted under shared resources). Hence, selecting new trials needs to be done carefully while understanding/forecasting the impact that this will have on other trials and the CRU resources.

These concepts motivate our planning approach to provide improved and controlled access, relative to FAS, for participants overall and especially those in high density trials. Under FAS, the participants enrolled in low density trials have very good access to care at the expense of the participants enrolled in high density trials. This is because the first available slot is always given to the current enrolling participant even if he/she is enrolled in a low density trial that is easy to book with low access (e.g., a short one visit trial), thus taking critical resource capacity away from the future participants in trial protocols that may be harder to schedule. In contrast, our CAPTAIN methodology reserves time slots (capacity) for first visits of each participant of each trial (see Sec. 3.2), while assuring that enough capacity is available for subsequent visits (Sec. 3.5).

1.3 CRU needs for capacity planning and scheduling tools: A structured survey

To clearly identify the status of common sense approaches in CRUs, MCRU conducted a survey across 80 NIH-sponsored Cancer Centers, with 34 responses. The most difficult challenge was identified as knowing how to “Schedule the number/variety of protocols requested,” with more than 95% selecting Strongly Agree or Agree. More than 80% of respondents agreed that a scheduling tool for predicting capacity and overutilization could help improve metrics. CRUs agreed or strongly agreed that “In an effective Clinical Research Unit Planning Tool, I would value the ability to: “predict the impact of a new trial (> 90%), reduce overutilization expenses (80%), and reduce time-to-first-visit (80%). A real time scheduling tool which confirms that the required resources are available for all visits at time of booking was considered very valuable or valuable by 92% of the centers.

1.4 Overview

The rest of this paper is organized as follows: Next, we review the related literature. Our CAPTAIN methodology is presented in Sec. 3. In Sec. 3.1 we discuss scientific value and the objective function. Sec. 3.2 defines the primary decision variables and characterizes the

dynamics of the stochastic booking process, called the BRPlan. Despite the fact that the system dynamics are non-linear in the decision variables (Eq. (1)), we show that we can capture the desired metrics exactly with linear constraints in a deterministic Mixed Integer Program (MIP) (Theorems 3.1-3.3). To keep practical applications tractable however, we develop a novel approximation that aggregates information of the historical booking process (Proposition 3.1). Sec. 3.3 details how we are able to constrain TFAV within the MIP (Theorem 3.2), Sec. 3.4 details the resources requirement constraints, and Sec. 3.5 characterizes the offered daily workload (Theorem 3.3). The online Appendix contains a list of notation, the proofs, and the complete math programming formulation. To help better understand the methodology and provide intuition, we numerically investigate the smaller, simpler example in Appendix A. Since we propose approximations, we need to validate our metrics. Our computations are validated in Appendix G, using a larger test suite with protocol characteristics from MCRU that are believed to be representative of many other sites. Sec. 4 discusses optimization results and demonstrates the capabilities and potential of CAPTAIN over the current FAS approach. Sec. 5 concludes the paper.

2 Literature survey

Rigorous research as well as the software vendor community have left the topic of CRU operations almost untouched. Vendors have targeted the support of individual trials without providing support for managing the CRU as a system of shared resources. Systems like EPICTM (see Anderson and Matessa 1997) provide basic scheduling and information management, but our research seeks to harness mathematical system modeling and optimization to provide functionality for planning, scheduling, and improved trial performance. Survey results (provided in Sec. 1.3) from many leading CRUs indicate that they are experiencing major difficulties and are eager for science-based engineering systems for planning and scheduling. Most of the literature in clinical trial management focuses on how to design trials (the protocols) that are successful and efficient throughout the different stages that lead to drug development/commercialization (see Varma et al. 2007, Felli et al. 2007). Some literature treats the recruiting process of participants (see McGarvey et al. 2007). At the same time, “patient recruitment should not start until the clinics and data center have demonstrated that they are properly staffed and equipped to support this activity,” states Meinert and Tonascia (1986).

The research models and algorithmic solutions in the literature do not provide a methodology that simultaneously accounts for the following elements that we integrate: (i) timetables for multi-visit care pathways of different durations and resource requirements with specified tolerances for their timing, (ii) a multi-month transient planning horizon, (iii) various room resources, (iv) staff with specific skills/capabilities, (v) variable and uncertain participant enrollments, (vi) estimating and controlling patient access by type, (vii) trial portfolio selection, (viii) estimating and controlling overutilization costs, (ix) approximately optimized booking methodologies, and (x) coordinating the nursing staff and the physical resources with participants’ treatment needs. Despite a large literature on scheduling and planning (e.g., Gupta and Denton 2008), there is little extant work to suit the combination of attributes of the problem at hand.

Gupta and Denton (2008) uses the term “indirect waiting” to describe the wait to get an appointment which we have captured at the daily level as TFAV. A prior model by Grunow et al. (2004) determines the best start times of trials (within a day) and the allocation of individual personnel to clinical tasks using an MIP model. But this paper does not consider items (ii), (iii), and (vi)-(x). Moreover, the trial protocols they considered do not require more than one day to complete, whereas the sequence of visits to complete a trial can span many weeks in our model. By using historical data to construct a matrix that assigns probabilities for patients’ ward locations overtime, Helm and Van Oyen (2014) presents a new method to predict and optimize the daily census at every ward in a hospital. By rearranging the elective/surgical schedules, they can control the offered load on each hospital ward in an equilibrium model that is cyclo-stationary. Our model differs from theirs in items (i), (ii), (iv), (vi), (vii), (viii), and (x). Relevant literature also includes scheduling outpatients to reduce waiting times such as Cayirli and Veral (2003) (see also the survey by Ahmadi-Javid et al. (2017)), and many others such as Phillips et al. (1995), Pinedo (2008), and Asmundsson et al. (2009).

In general, there is a significant amount of work in patient scheduling. Gocgun and Puterman (2014) does a good job of categorizing the patient scheduling literature into either *allocation scheduling* or *advanced scheduling*, see the references therein. While for allocation scheduling patient requests are either rejected or serviced immediately, advanced scheduling involves the booking of services over a future horizon. Ours is a decision model with elements of both advance scheduling (in terms of capacity reservation) and allocation scheduling (in terms of acceptance/rejection of candidate trials). The prior literature does not treat the distinguishing features of CRUs, such as choices of which trials to accept, the enrollment targets for trials, the complex patient pathways, and multidimensional capacitated resources. Our work though is not the first to model settings in which each patient (job) receives a series of visits spanning multiple days and having deadlines or time windows for subsequent visits (see Gocgun and Ghate 2012, Sauré et al. 2012, Turkan et al. 2012, Hulshof et al. 2013, Gocgun and Puterman 2014, Condotta and Shakhlevich 2014, Sauré et al. 2019).

Some of the more closely related allocation scheduling works are in the area of chemotherapy appointment booking, including those of Turkan et al. (2012), Gocgun and Puterman (2014), and Condotta and Shakhlevich (2014). Turkan et al. (2012) considers a finite, deterministic population of chemotherapy patients with multiple visits over time and formulate an MIP with an objective that minimizes patients’ wait time from their earliest start dates. As many patients as possible are fit into the planning horizon while also minimizing overutilization. Similar to our work, that one solves a scheduling problem with limited clinical resources, including beds/chairs, nurses, and pharmacists. In contrast, we model participant arrivals within a trial as stochastic processes, while constraining population average TFAVs and overutilization. Stochastic arrivals induce dynamics similar to those of queueing networks with blocking, which are not easy to analyze except for special cases. For tractability in practical applications, we rely on linearization of our metrics with respect to our decision variables through novel approximations. We also include the practical complication of time windows. Gocgun and Puterman (2014) study an ADP model similar to that of Patrick et al. (2008), which includes stochastic arrivals, and

use approximate dynamic programming (ADP) to consider different patient types that require different levels of access to a single appointment for a diagnostic resource. Our system differs in items (ii)-(iv), (vii), and (x). Somewhat similar to the time windows that we consider, Gocgun and Puterman (2014) allows for an interval to be specified (earliest date and latest date) for a chemotherapy appointment, and penalizes appointments that are either early or late. Sauré et al. (2019) also builds upon Patrick et al. (2008), using ADP. Like our problem, that one is an advance scheduling problem with multiple classes of patients with stochastic arrivals, heterogeneous service times, and resource capacity, accounting for access delays (with penalties for missing the target) and overtime (plus idle time). Unlike ours, that problem assumes strict priority sub-classes within patient classes corresponding to patient urgency. That paper presents a closed form solution for the deterministic service time case. Interestingly, for its setting, that study suggests that not much can be gained from modeling stochastic service times. Condotta and Shakhlevich (2014) presents an MDP model for both intra-day and multi-day appointment patterns. Two of the differences between our work and these is that we are concerned with making decisions at the planning level, not necessarily at the execution level, and these other works do not address the portfolio selection problem that we are facing. Our main modelling contribution is to linearize this complex system using good approximations at a level of accuracy that is useful at the planning level. We capture queueing delays without resorting to a very detailed model of individual requests in order to maintain tractability for the size of problems that are faced by operational planners of a CRU.

Diamant et al. (2018) uses ADP to study an outpatient care program in which patients receive multiple visits and then have surgery, which has a stochastic duration. Having stochastic visit durations and a complex model of patient no-shows are elements we do not model; furthermore, they determine which treatments are provided on the visit day (or be contacted at a later date with their next appointment) based on the arrivals and no-shows. Unlike that paper, which is non-committal when promising an appointment, we focus on reducing access delays through a variety of constraints in a committal setting.

Our work is an advance scheduling model that relies upon the notion of capacity reservation/allocation. That notion is found in the revenue management literature (e.g., Akkan 1997, Hsu and Wang 2001, Mula et al. 2006, Talluri and Van Ryzin 2006, Gupta and Wang 2008), but that literature has not addressed the type of system seen in CRUs.

It should be mentioned that there has been work on the optimization of adaptive clinical trials, including Berry (1978) and Ahuja and Birge (2016). We do not consider adaptive clinical trials in our model. In our model (and in practice at many CRUs including MCRU), the initial visit of a protocol starts a sequence of nearly deterministic visits, within the relatively short time windows. Our model, at least in its current state then, does not apply to adaptive trials.

Some of the ideas of our model though have recently been extended to another application area – patient access for a network of outpatient services, in Deglise-Favre-Hawkinson (2015) and Deglise-Hawkinson et al. (2018). In the latter, the workload of a patient who is granted access to the network induces a downstream workload that is stochastic because an

initial visit may, probabilistically, result in subsequent visits (perhaps in another department) depending on the results of diagnostic tests, etc. That model does not capture nearly the same level of complexity in the resource requirements including specific nurses with specific skills, rooms, equipment, etc.

3 The CAPTAIN Methodology

CAPTAIN solves three interrelated problems: (i) selection of the portfolio of candidate clinical trials that can and should be added to maximize the scientific/financial value extracted, (ii) creation of the daily BRPlan (Booking Reservation Plan), which is a capacity allocation plan that reserves slots for trial visits, and (iii) construction of a daily nurse schedule (NSPlan) that accounts for every planning visit's requirements for nurse skills.

One contribution of this paper is to advance the notion that, while it is desirable to conduct a new candidate trial, it may be best to reject the trial (or delay it until the system congestion lowers as trials complete). In a setting where some trials must be rejected, adding a "scientific value gained" (SVG) to each trial allows us to prioritize the acceptance of some trials over the others, while recognizing that this decision must be done jointly with the capacity profile demanded by each trial. To facilitate item (i), we will use a scalar metric of SVG and incorporate it as one component in the decision of which trials to conduct (see more details on the generation of this metric in Appendix B). If the CRU/model must accept one or more new trials to be added, the optimization would still have very significant benefits by virtue of setting the capacity reservation values according to the TFAV targets.

Our goal is to construct a finite horizon optimized transient BRPlan and NSPlan that are resolved in a type of rolling horizon manner. This optimization model will estimate and control the critical CRU metrics defined earlier. An equilibrium model does not apply because the portfolio of active trials in a CRU tend to vary with time: some new trials can be added, while some will end or be interrupted. The length of this planning horizon, for purposes of modeling the arrival process, will be D business days of participant enrollments (e.g., 60 days), but with regards to reserving capacity the horizon is extended until the latest date at which arrivals on day D can be scheduled and meet their TFAV constraints. In a rolling horizon framework, the planning horizon can be ended and the problem restarted with proper transfer of information to generate the proper initial condition if new trials must be added or one terminates. For ease of presentation, we do not explicitly detail the implementation of the rolling horizon updates, but Appendix C gives a description.

To support implementation, it is important that we be realistic in modeling the way that CRUs would transition from any prior method for scheduling patients to our methodology. This requires an initial condition for the visits already booked using the existing paradigm. In our numerical analysis we simulate the current active portfolio in the CRU, $\mathcal{P}^0$, so that the initial condition is populated with appointments for T days (e.g., 720 business days). This initial condition induces a certain "committed" workload/capacity for each resource (skills and rooms) on each day of our planning horizon. In our study, this is generated by the FAS scheduling of patients that enrolled prior to CAPTAIN's period of planning. Those initial workload conditions will be taken as inputs to our math program. Our question is: how

should the resource capacity still available be allocated to schedule our new participant enrollments from both $\mathcal{P}^0$ and an optimized set of candidate trials, while forecasting and controlling the concerned metrics? Note that all mathematical notation is described in Appendix D.

3.1 Scientific value and the objective function

A key objective of a CRU is to extract the most value from the resources available. To help evaluate potential trials to be added, the model accepts an input SVG M_k for every candidate clinical trial $k \in \mathcal{P}^{cand}$. This is an obvious mechanism to obtain the greatest good from the CRU resources. Of course, setting $M_k = 1$ for all trials k is still a very useful model, provided constraints are placed on TFAV to provide equal access to all despite logistical differences between trials. In current practice, any new participant enrollment is scheduled to the next available slot such that all future visits can adhere to the protocol requirements while also respecting capacity constraints. As a side effect, this results in a system which offers the best access (low TFAV) to the participants in the low density trials. Complex trial protocols suffer with long TFAV metrics, which is a barrier to enrollment, may undermine the participant's health care, and lengthens the duration of the trial. For our methodology, scientific value is a suitable objective function for allocating capacity, while satisfying constraints on metrics that include TFAV and overutilization limits. Scientific value can have different meanings for different CRUs. While scoring M_k is a challenge outside the scope of this paper, one view is to rate the trials based on the scientific/research knowledge that can be gained from them (which is already done in practice). Scientific value could also incorporate the financial value gained from a trial, especially for industry funded trials. It is not difficult to set up the scoring system to ensure that the scientific goals take priority over finances by weighting the various categories appropriately. Additional information on scientific value can be found in Appendix E.

One possible objective of the MIP can be summarized neatly by maximizing the following function, but the deep computations all happen in the constraints, which are provided later:

$$\sum_{k \in \mathcal{P}} M_k h_k,$$

where h_k is a binary decision variable set to 1 if and only if trial $k \in \mathcal{P}^{cand}$ will be activated, and $\mathcal{P} \equiv \mathcal{P}^{cand} \cup \mathcal{P}^0$. Note that h_k will be a parameter equal to 1 for trials $k \in \mathcal{P}^0$ that are still ongoing at the beginning of our planning horizon. The following sections will introduce and justify the modeling of our system's constraints, and a complete version of the MIP used for our case studies is in Appendix F.

3.2 The Booking Reservation Plan (BRPlan)

The BRPlan is a patient admission control mechanism that packs the participants' first visit requests into the planning space of physical, human, and time resources via planned appointment time blocks for each type of first visit planned and accounting for all follow-up visits. Our decision variables, denoted $\Theta_{k,d}$ will set a maximum limit on the number of type

k participants that can start their treatment sequence on day d of our planning horizon, thereby controlling the participant flow of enrollments. Our planning model is at the daily level, so the model does not specifically sequence the visits within a day (a complication that remains for future research). For each active and candidate trial $k \in \mathcal{P}$, these BRPlan decision variables $\Theta_{k,d}$ will take values in $\mathbb{Z}^+$, and are defined over the set of days $\{1, \dots, D + \overline{TFAV}_k^{max}\}$. The horizon extends $\overline{TFAV}_k^{max}$ days beyond the end of the planning horizon, D , because CAPTAIN schedules the enrolling participants (up to day D) within the population maximum TFAV, $\overline{TFAV}_k^{max}$, which is set by the decision maker. To compute the TFAV access metric and to optimize the BRPlan, a given type k participant enrolling will always be given the first available type k reserved capacity slot within the BRPlan. After the first visit, we account for participant's preference within the time windows of subsequent visits. The first visit for type k is booked to the first available day with sufficient reserved capacity by the BRPlan, so that we can emphasize the effectiveness of the capacity management method of a CRU to guarantee a level of access to the most urgent patients and to do so without dependence on the participants' personal preferences. The BRPlan differs from the current FAS methodology, because it selects the first available slot reserved for type k participants (which was determined by an informed decision of all protocols' needs derived via optimization).

To quantify the stochastic booking process, we assume that a type k participant enrolling on day t can be scheduled for their first visit no earlier than day $t + 1$. We provide the following key definition:

Definition 3.1. *Let $\alpha_{k,d,t}$ denote the nonnegative integer valued random number of type k first visit participants enrolled on day $t \in \{0, \dots, D\}$ and booked under the BRPlan on day $d \in \{t + 1, \dots, D + \overline{TFAV}_k^{max}\}$.*

One can compute $\alpha_{k,d,t}$ recursively while considering the type k daily demand/enrollment random variables denoted by $A_{k,t}$, $t \in \{0, \dots, D\}$, taking values in $\mathbb{Z}^+$, and the previously defined BRPlan type k admission limit decision variables, $\Theta_{k,d}$, as:

$$\alpha_{k,d,t} = \min \left\{ \left(\Theta_{k,d} - \sum_{l=0}^{t-1} \alpha_{k,d,l} \right), \left(A_{k,t} - \sum_{l=t+1}^{d-1} \alpha_{k,l,t} \right) \right\}. \quad (1)$$

In this expression, (i) the first term of the minimum is the integer number of type k reservations/slots remaining on day d for first visits after all participants who enrolled before day t have been scheduled, and (ii) the second term of the minimum is the random number of type k participants who enrolled on day t and are not booked for their first visit appointment before day d (or later). In the illustration in Fig. 3, the allocated capacity, Θ , is determined at time 0, before enrollments are realized (on the left). Day 0 enrollments are eligible to be booked on days $d > 0$ – not the same day. For booking (on the right) on day 1, we see $\alpha_{k,1,0} = A_{k,0}^{(w)}$ with w denoting a realization of the underlying probability space upon which the stochastic model rests, because term (ii) achieves the minimum since

$\Theta_{k,d} > A_{k,0}^{(w)} - \sum_{l=1}^{d-1} \alpha_{k,l,0}$. Also, $\alpha_{k,d,0} = 0$, $d > 1$, because all prior type k arrivals on day 0 were scheduled on day 1. We also see that on day 2 $\alpha_{k,2,1} = \Theta_{k,2} - \alpha_{k,2,0} = \Theta_{k,2}$, because (i) achieves the minimum since $\Theta_{k,2} < A_{k,1}^{(w)}$. We have $\alpha_{k,3,1} = A_{k,1} - \alpha_{k,2,1} = A_{k,1}^{(w)} - \Theta_{k,2}$ because day 3 has sufficient capacity to complete the type k arrivals on day 1 (implying that $\alpha_{k,d,1} = 0$ for $d > 3$ because $\Theta_{k,d} \geq (A_{k,1}^{(w)} - \sum_{l=1}^{d-1} \alpha_{k,l,1})$, $d > 2$).

The dynamics of our system are conceptually similar to the idea of blocking in queueing networks, with dynamics that are challenging to analyze except in special cases and that are non-linear in our main decision variables Θ . What we will do is translate the set of stochastic non-linear equations (1) into a set of deterministic constraints that are linear in Θ and capture the mean of the $\alpha_{k,d,t}$ random variables. This information is needed to accurately estimate (i) how demand will be fulfilled according to a given BRPlan Θ , (ii) the delay to obtain a first visit, and (iii) how the CRU's resources will be utilized.

We will show (in Lemma 3.1 and Theorem 3.1) that keeping full state information of prior enrollments allows us to compute our metrics exactly and linearly in our decision variables. However, an important mechanism to keep our optimization tractable for practical applications is to approximate the information state by aggregating the enrollment information for the m or more days prior to current day t , with $m \geq 1$. The random variables $A_{k,t}$ take values in the set $I = \{0, 1, \dots, j^*(t, m)\}$, with $j^*(t, m)$ being the maximum number of type k enrollments on a given day. The aggregation that we propose is defined as follows:

Definition 3.2. For all $k \in \mathcal{P}$, $t \in \{1, \dots, D\}$, $d \in \{t+1, \dots, D + \overline{TF\Delta V}_k^{max}\}$, and memory level $m \geq 1$, we define the set $\mathcal{A}_{k,d,t(m)}$ with elements $\vec{a}_{k,d,t}^{(t,m)} \in \mathbb{R}^+$, where $\vec{a}(t, m) = (a_t, a_{t-1}, \dots, a_{j^*(t,m)}, a)$, $j^*(t, m) = \max\{t-m+1, 1\}$, $a \in \{0, 1, \dots, j^*(t, m)\}$, and $(a_j)_{j \in \{j^*(t,m), \dots, t\}} \in I$. The elements $\vec{a}_{k,d,t}^{(t,m)}$ of the set $\mathcal{A}_{k,d,t(m)}$ represent the conditional expectation of the number of patients that enrolled on day t and are scheduled on day d given a partial history of prior enrollments $(A_{k,t}, \dots, A_{k,j^*(t,m)}, \sum_{i=0}^{j^*(t)-1} A_{k,i}) = (a_t, a_{t-1}, \dots, a_{j^*(t,m)}, a)$, where the oldest portion of the history, days $\{0, 1, \dots, j^*(t, m)-1\}$, are aggregated as a sum, while the most recent m days are retained individually for $t \geq m$.

In the above, $j^*(t, m)$ represents the number of days over which the total/aggregate number of enrollments will serve as an “information state.” We aggregate from day 0 to $j^*(t, m)-1$ in our planning horizon. For each single day from $j^*(t, m)$ to t , we will include the daily enrollment information in the state variables of the model. As an example, let's assume that t is day 11 and $m = 5$ as illustrated in the left-hand side of Fig. 4. Observe that $j^*(11, 5) = 7$ and that set $\mathcal{A}_{k,d,11(5)}$ will contain the conditional expectations $\vec{a}_{k,d,11}^{(11,5)}$, $\vec{a}(11, 5) = (a_{11}, a_{10}, \dots, a_7, a)$, of the number of participants enrolled on day 11 and scheduled on day $d > 11$ given the following partial history of enrollment: (a) the daily enrollment information from the previous $m=5$ days, days 7 to 11 ($A_{k,i} = a_i, \forall a_i \in I, \forall i \in \{7, \dots, 11\}$),

and (b) the aggregated prior information of total enrollments from days 0 to 6 ($\sum_{i=0}^6 A_{k,i} = a$, $\forall a \in \{0, \dots, 7\}$). The same concept applies for all $1 \leq m \leq t$. Represented on the right-hand side of Fig. 4 is the case when $m = t$. For example, for $m = 20$, by definition $j^*(11, 20) = 1$, and we notice that set $\mathcal{A}_{k,d,11}(20)$ will contain the conditional expectations $\vec{\alpha}_{k,d,11}^{\vec{a}(11,20)}$, $\vec{a}(11, 20) = (a_{11}, a_{10}, \dots, a_1, a)$, of number of participants enrolled on day 11 and scheduled on day $d > 11$ given a full daily history of enrollment from days 0 to 15 (where day 0's enrollment realization will be characterized by a).

Remark 3.1. In Definition 3.2, we omitted the first day where $t = 0$ due to notational complexity. To address this issue, we will assume $j^(0, m) = 0 \forall m \geq 1$ and use the convention $\vec{a}(0, m) := a_0$, to express the elements of the set $\mathcal{A}_{k,d,0}(m)$ similarly to Definition 3.2. Hence, in the case $t = 0$, $\vec{\alpha}_{k,d,0}^{\vec{a}(0,m)} \forall m \geq 1$ will represent the conditional expectation of the number of patients that enrolled on day 0 and are scheduled on day d given that there were $A_{k,0} = a_0$ type k enrollments on day 0.*

In Lemma 3.1 and Theorem 3.1, we show that a linear MIP can be formulated in our decision variables when we keep a full information state of prior enrollments: $m = \infty$.

Lemma 3.1. For all $k \in \mathcal{P}$, $t \in \{0, \dots, D\}$, $d \in \{t + 1, \dots, D + \overline{TFAV}_k^{max}\}$, the elements $\vec{\alpha}_{k,d,t}^{\vec{a}(t,\infty)} \in \mathcal{A}_{k,d,t}(\infty)$ can be expressed in an MIP as a set of linear constraints that are linear in our decision variables Θ .

The proofs of all lemmas, propositions, and theorems are in Appendix B. We allow general (and time dependent) distributions for type k demand, so the model can use historical/empirical distributions from the given (or from a similar) trial's enrollment processes or use expert opinion to account for different rates of enrollments with time, which may be useful during periods of trial advertisement. With a finite I , we let $P_{k,t}(a_t)$, $a_t \in I$ with $I \equiv \{0, 1, \dots, \bar{I}\} \subset \mathbb{Z}^+$, be the probability that there are a_t enrollments of type k on day t (which can be drawn from historical and protocol data). Finally, we make the assumption that A_{k,t_1} is independent of A_{k,t_2} for all $t_1 \neq t_2$, which is a fair assumption in practice even when the arrival process is nonstationary over time due to promotions.

Theorem 3.1. Given any BRPlan matrix Θ , and as $m \rightarrow \infty$ in the case of unbounded support for the enrollment distribution, the expected value of $\vec{\alpha}_{k,d,t}$, $\bar{\alpha}_{k,d,t}$, can be computed linearly in our decision variables Θ as:

$$\bar{\alpha}_{k,d,t} = \sum_{\vec{a}(t,\infty) \in \mathcal{A}_{k,d,t}(\infty)} \vec{\alpha}_{k,d,t}^{\vec{a}(t,\infty)} \cdot \prod_{l=0}^t P_{k,l}(a_l). \quad (2)$$

For every $k \in \mathcal{P}$, $t \in \{0, 1, \dots, D\}$, and d such that $t < d \leq D + \overline{TFAV}_k^{max}$, the set $\mathcal{A}_{k,d,t}(\infty)$ will be composed of $(+1)^{t+1}$ elements. In our setting, clinical trial daily enrollment/demand rates

are rather low (generally less than 1 enrollment per day), which allows us to choose a small while not losing much accuracy. We run into dimensionality issues as the length of our planning horizon gets large (e.g., t is 60 for a 12 business week planning horizon). For tractability, we approximate our booking methodology by reducing the cardinality of the set $\mathcal{A}_{k,d,t}(\infty)$. In all that follows, we use the set defined as $\mathcal{A}_{k,d,t}(m)$ in Definition 3.2, with a partial aggregation of the enrollment information from days 0 to $j^*(t,m)-1$.

The goal is to express $\vec{\alpha}_{k,d,t}^{(t,m)} \in \mathcal{A}_{k,d,t}(m)$ linearly in our decision variables, but Eq. (1) requires taking the conditional expectation of the minimum of two random variables. In the case $m = \infty$ (see Lemma 3.1), conditioning on the complete history of enrollments allows the two elements of the minimum of Eq. (1) to be deterministic, so the expected value of the minimum will be the minimum of the expected values of each term. In the case $m < t$ however, this won't hold, and we rely on the following approximation. We define $\mathcal{A}_{k,d,t}(m)$ as the set with elements $\hat{\vec{a}}_{k,d,t}^{(t,m)}$, to approximate $\mathcal{A}_{k,d,t}(m)$. So, we define $\hat{\vec{a}}_{k,d,t}^{(t,m)}$ to be:

$$\begin{aligned} & \min \left\{ \mathbb{E} \left[A_{k,t} - \sum_{l=t+1}^{d-1} \alpha_{k,l,t} \left(A_{k,t}, A_{k,t-1}, \dots, A_{k,j^*(t,m)}, \sum_{i=0}^{j^*(t,m)-1} A_{k,i} \right) \right] \right. \\ & \quad \left. = (a_t, a_{t-1}, \dots, a_{j^*(t,m)}, a) \right\}, \\ & \mathbb{E} \left[\Theta_{k,d} - \sum_{l=0}^{t-1} \alpha_{k,d,l} \left(A_{k,t}, A_{k,t-1}, \dots, A_{k,j^*(t,m)}, \sum_{i=0}^{j^*(t,m)-1} A_{k,i} \right) \right] \\ & \quad \left. = (a_t, a_{t-1}, \dots, a_{j^*(t,m)}, a) \right\}. \end{aligned} \quad (3)$$

We will assume that the elements $\vec{\alpha}_{k,d,t}^{(t,m)}$ of $\mathcal{A}_{k,d,t}(m)$ are close to $\hat{\vec{a}}_{k,d,t}^{(t,m)} \in \mathcal{A}_{k,d,t}(m)$ for all $k \in \mathcal{P}$, $\forall t \in \{0, \dots, D\}$, $\forall d \in \{t+1, \dots, D + \overline{TF}AV_k^{max}\}$. In other words, we assume equality in the well known Jensen's inequality (i.e., we assume $\mathbb{E}[\min\{X, Y\} | Z] = \min\{\mathbb{E}[X | Z], \mathbb{E}[Y | Z]\}$ for the random variables X , Y , and Z). The less variability there is in the random variables $\mathbb{E}[X | Z]$ and $\mathbb{E}[Y | Z]$, the more accuracy we will have in our approximation. Hence, the more enrollment information we condition $\alpha_{k,d,t}$ on (i.e., the larger m is), the closer $\hat{\vec{a}}_{k,d,t}^{(t,m)}$ will be to $\vec{\alpha}_{k,d,t}^{(t,m)}$. The following result is used to approximate mean offered workload (Theorem 3.3).

Proposition 3.1. *Given $\hat{\vec{a}}_{k,d,t}^{(t,m)} \in \mathcal{A}_{k,d,t}(m)$ defined in (3), and any BRPlan matrix Θ , as $\bar{I} \rightarrow \infty$, $\bar{\alpha}_{k,d,t}$, the expected value of $\alpha_{k,d,t}$, can be approximated linearly in our decision variables Θ as:*

$$\bar{a}_{k,d,t} \approx \sum_{\substack{\vec{a}_{k,d,t}^{(t,m)} \\ \vec{a}_{k,d,t}^{(t,m)} \in \mathcal{A}_{k,d,t(m)}}} \hat{a}_{k,d,t}^{(t,m)} \cdot \left(\prod_{l=j^*(t,m)}^t P_{k,l}(a_l) \right) \cdot \bar{P}_{k,j^*(t,m)}(a). \quad (4)$$

where $\bar{P}_{k,j^*(t,m)}(a) = \sum a_0 + a_1 + \dots + a_{j^*(t,m)-1} = a \prod_{l=0}^{j^*(t,m)-1} P_{k,l}(a_l)$ is the probability of having exactly a type k enrollments in the first $j^*(t,m)$ days (from day 0 to $j^*(t,m)-1$) of our planning horizon.

Again, the approximation in (4) is due to (i) the aggregation of the historical booking information and (ii) the replacement of Jensen's inequality with an equality. Note that as m decreases, the cardinality $|\mathcal{A}_{k,d,t(m)}|$ decreases while trading off some accuracy when computing $\bar{a}_{k,d,t}$ (since we aggregated information).

3.3 Formulation and control of the Time to First Available Visit (TFAV)

We now consider the control of participants' access to trials. We formulate three TFAV metrics for a given trial k : (i) the average TFAV, denoted $TFAV_k^{avg}$, which is the mean TFAV averaged over all participants of type k that enroll within our time horizon, (ii) the TFAV-exception rate, denoted $TFAV_{k,t}^{except}(\omega)$, which is the expected value of the percentage of type k participants that enroll on day t and have to wait strictly longer than ω days before their initial visit, and (iii) the population maximum TFAV, denoted $\overline{TFAV}_k^{max}$, which captures the maximum expected number of days that a type k participant have to wait within our time horizon. In (ii), the TFAV-exception rate will be defined for each day $t \in \{0, \dots, D\}$ by a type k sequence $(TFAV_{k,t}^{except}(\omega))_{\omega \in \mathcal{T}, \mathcal{T} \subset \{0, 1, 2, \dots, \overline{TFAV}_k^{max}\}}$, where each element is the expected fraction of type k participants enrolling on day t that will exceed ω days of waiting for their first visit (i.e., won't be scheduled before $t + \omega + 1$). By allowing the CRU administrator to control (i)-(iii), we will be able to reduce the access variability between participants of a given trial and shape the access delays across trials. The elements and the cardinality of the set $\mathcal{T}$ are to be custom selected to meet the relative trial priorities and participant treatment needs.

Theorem 3.2. *The average type k TFAV can be expressed linearly in our BRPlan variables Θ as follows:*

$$TFAV_k^{avg} = \sum_{t=0}^D \sum_{d=t+1}^{D + \overline{TFAV}_k^{max}} \sum_{\substack{\vec{a}_{k,d,t}^{(t,\infty)} \\ \vec{a}_{k,d,t}^{(t,\infty)} \in \mathcal{A}_{k,d,t(\infty)}}} (d-t) \cdot \hat{a}_{k,d,t}^{(t,\infty)} \cdot \frac{\prod_{l=0}^D P_{k,l}(a_l)}{\sum_{i=0}^D a_i} \quad (5)$$

Next, the customer-average type k TFAV exception rate on day t for a chosen delay ω is computed linearly in Θ as follows:

$$TFAV_{k,t}^{except}(\omega) = \mathbb{E} \left[\sum_{d=t+\omega+1}^{D+\overline{TFAV}_k^{max}} \frac{\alpha_{k,d,t}}{A_{k,t}} \right] \quad (6)$$

$$= \sum_{(a_t, a_{t-1}, \dots, a_0) \in I^{t+1}} \sum_{d=t+\omega+1}^{D+\overline{TFAV}_k^{max}} \mathbb{E} \left[\frac{\alpha_{k,d,t}}{A_{k,t}} \right] \quad (7)$$

$$\begin{aligned} & \left[(A_{k,t}, A_{k,t-1}, \dots, A_{k,0}) = (a_t, a_{t-1}, \dots, a_0) \right] \cdot \prod_{l=0}^t P_{k,l}(a_l) \\ &= \sum_{d=t+\omega+1}^{D+\overline{TFAV}_k^{max}} \sum_{\substack{\vec{a}(t,\infty) \\ \alpha_{k,d,t} \in \mathcal{A}_{k,d,t}(\infty)}} \frac{\alpha_{k,d,t}(\vec{a}(t,\infty))}{a_t} \cdot \prod_{l=0}^t P_{k,l}(a_l). \end{aligned} \quad (8)$$

Eq. (8) can be derived by conditioning Eq. (6)'s expectation with respect to $(A_t, \dots, A_0) = (a_t, \dots, a_0)$. Assuming a finite set I (which is a good assumption in practice), we can change the order of summation, and by using the definition of set $\mathcal{A}_{k,d,t}(\infty)$ and its elements $\vec{a}(t,\infty)$, we can derive Eq. (8). It will be linear in our decision variables thanks to our result in Lemma 3.1.

Remark 3.2. Note that when solving this problem for long planning horizons, we can approximate the average TFAV (resp., TFAV exception rate) formulation in Theorem 3.2 (resp., Eq. (8)) by computing $TFAV_k^{avg}$ (resp., $TFAV_{k,t}^{except}$) as a linear combination of $\hat{\alpha}_{k,d,t}^{\vec{a}(t,m)}$ terms (see (3)), as follows:

$$\begin{aligned} TFAV_k^{avg} &\approx \sum_{t=0}^D \sum_{d=t+1}^{D+\overline{TFAV}_k^{max}} \sum_{\substack{\vec{a}(t,m) \\ \hat{\alpha}_{k,d,t} \in \mathcal{A}_{k,d,t}(m)}} (d-t) \cdot \hat{\alpha}_{k,d,t}^{\vec{a}(t,m)} \\ &\cdot \sum_{(a_D, a_{D-1}, \dots, a_{t+1}) \in I^{D-t}} \frac{\bar{P}_{k,j^*(t,m)}(a) \cdot \prod_{l=j^*(t,m)}^D P_{k,l}(a_l)}{a + \sum_{l=j^*(t,m)}^D a_l}, \end{aligned} \quad (9)$$

$$\begin{aligned} TFAV_{k,t}^{except}(\omega) &\approx \sum_{d=t+\omega+1}^{D+\overline{TFAV}_k^{max}} \sum_{\substack{\vec{a}(t,m) \\ \hat{\alpha}_{k,d,t} \in \mathcal{A}_{k,d,t}(m)}} \frac{\hat{\alpha}_{k,d,t}^{\vec{a}(t,m)}}{a_t} \cdot \bar{P}_{k,j^*(t,m)}(a) \\ &\cdot \prod_{l=j^*(t,m)}^t P_{k,l}(a_l). \end{aligned} \quad (10)$$

We can limit $TFAV_k^{avg}$ by constraining it by a specified upper bound on average access wait, denoted $\overline{TFAV}_k^{avg}$. Also, we let $p_k(\omega)$, with $\omega \in \mathcal{T}$, be the limit placed on the expected percentage of type k participants that enrolled on day t that are permitted ω days of waiting for their first visit. Note that setting $p_{k,t}(\overline{TFAV}_k^{max}) = 0$ for all $t \in \{0, \dots, D\}$ is equivalent to setting a population maximum TFAV limit of $\overline{TFAV}_k^{max}$ on trial k . This assures that type k enrollments on any given day t will be booked for their first visit on a day $d \leq t + \overline{TFAV}_k^{max}$.

3.4 The stochastic Participant Resource Requirement Model (PRRM)

The PRRM is a stochastic location process model (a particular random field) we developed to tractably capture randomness in participant resource needs over time. It was inspired by Helm and Van Oyen (2014). Let the set of all possible physical locations/resources be $\mathcal{R}$ and the set of all skills of the personnel be $\mathcal{S}$. Then the entire set of resources is indexed by r in $\mathcal{R} \cup \mathcal{S}$. Note that specific skills are required on each visit of a trial, rather than specific nurses. Each skill is mapped via the nurse skill sets to one or more staff possessing that skill. This captures the unique skills of each individual staff member. First, we focus on the workloads induced on skills, and later we will assign those workloads to specific nurses.

The protocol information for any trial k includes (i) the number of visits V_k in the trial, (ii) the time window requirements on each visit v , $\Omega_{k,v}$, which represents a finite set of days that visit v can be scheduled after the first visit without provoking a deviation from protocol (patient preference is exercised only within those time windows), (iii) the set of clinical resources needed (for example it could be skill 1, skill 5, and the procedure room) needed for each visit v , and (iv) the deterministic workload, $\bar{w}_{k,v,r}$, induced by a visit v on resource r . As an example of item (ii), if $\Omega_{k,v} = \{10, 11, 12\}$ then a type k participant will require resources on a given day between 10 and 12 days after their first visit. There may exist some protocols where visit v has no time window requirement, and in this case we let $\Omega_{k,v} = \{1, 2, \dots, L^{max}\}$, where L^{max} is the maximum length (in days) of the longest protocol (for a single participant). For first visits, there are no time windows since the participant can start on any day with available capacity, so by convention, we let $\Omega_{k,1} = \{0\}$, $\forall k \in \mathcal{P}$. We use protocol information in items (i)-(iv) to create a participant flow model that forecasts the required workload of participants of type k , on and after their first visit. The vector $PRRM_{k,d_1}(d)$ determines, for a randomly sampled participant starting protocol k on day d_1 , the number of hours he or she will need each resource $r \in \mathcal{R} \cup \mathcal{S} \cup \mathcal{H}$ on day d , where $\mathcal{H}$ is the “home” resource – the participant has no appointment in the CRU. Letting e_r be the unit vector with a 1 in the r^{th} column, our goal is to compute $\mathbb{P}(PRRM_{k,d_1}(d) \cdot e_r = w)$, the probability that a participant starting protocol k on day d_1 will require w hours of resource r on day d . We will assume, in this paper, that participants enrolling in an active trial k on day d_1 have a uniform distribution of preferred days within their visits’ time windows, which generates the following probability distribution of the workload induced on resource r (skills and rooms) on day d .

$$\mathbb{P}(\text{PRRM}_{k,d_1}(d) \cdot e_r = w) = \sum_{v=1}^{V_k} 1_w = \bar{w}_{k,v,r} \frac{1_{\{(d-d_1) \in \Omega_{k,v}\}}}{|\Omega_{k,v}|}. \quad (11)$$

Remark 3.3. From Eq. (11), we can compute the expected value of the PRRM, denoted $\overline{\text{PRRM}}_{k,d_1,r}(d)$, for all trials k :

$$\overline{\text{PRRM}}_{k,d_1,r}(d) = \mathbb{E}(\text{PRRM}_{k,d_1}(d) \cdot e_r) = \sum_{v=1}^{V_k} \frac{1_{\{(d-d_1) \in \Omega_{k,v}\}}}{|\Omega_{k,v}|} \bar{w}_{k,v,r}.$$

While here we assume a uniform distribution within the days of the visit time windows to model patient preferences, we note that in practice this can be generalized by using patient preference estimated from historical distributions. Also, the required time window of visit v of a protocol is only dictated by the timing of the first visit. A trivial extension of our PRRM formulation would be to consider the time dependence between two subsequent visits (e.g., visit 3 has to occur between 10 and 12 days after visit 1, but also exactly one day after visit 2).

3.5 Formulation of the offered daily workload

As is common in queueing problems, computing the workloads in the system is fundamental. The daily offered workload for our planning horizon will depend on (i) the existing workload on resource r and day d induced by participants that enrolled prior to the start of our horizon, denoted $W_r^{\text{init}}(d)$, (ii) the portfolio of the selected candidate trials chosen at the beginning of our planning horizon, (iii) the number of daily participant bookings for their first visit by trial, to be optimized by the reserved resource capacity plan, BRPlan, and (iv) the PRRM model of patient flow through subsequent visits. Since we are considering enrollments from days 0 to D , and we limit the trial k population maximum TFAV level to $\overline{\text{TFAV}}_k^{\text{max}}$, we need to predict the offered workloads up to day $D + L^{\text{max}} + \max_k \{\overline{\text{TFAV}}_k^{\text{max}}\}$ (where again L^{max} is the duration in days from the first visit to the last visit of the longest protocol). Let $\mu_{k,d}$ be the random variable designating the number of type k participants scheduled on day d of our planning horizon under the BRPlan policy, and $\bar{\mu}_{k,d}$ be its mean. The following theorem shows how to compute the offered workload on any day d , for any resource $r \in \mathcal{R} \cup \mathcal{S}$.

Theorem 3.3. Given $\bar{\mu}_{k,d_1}$, the mean offered workload for resource $r (r \in \mathcal{R} \cup \mathcal{S})$ on day d of our planning horizon $\bar{W}_r(d)$, for $d = 1, 2, \dots, D + L^{\text{max}} + \max_k \text{TFAV}_k^{\text{max}}$, can be expressed linearly as follows:

$$\bar{W}_r(d) = W_r^{init}(d) + \sum_{k \in \mathcal{S}} \frac{\min\{d, D + \overline{TFAV}_k^{max}\}}{d_1 = 1} \bar{\mu}_{k, d_1} \cdot \overline{PRRM}_{k, d_1, r(d)},$$

$$\bar{\mu}_{k, d} = \mathbb{E}[\mu_{k, d}] = \sum_{t=0}^{d-1} \bar{\alpha}_{k, d, t}.$$

While Theorem 3.3 applies to both the rooms and skills of the personnel, we now focus on the human resources, which we refer to as nurse resources (as non-nursing staff can be treated similarly). Below, index n denotes the n^{th} nurse in set $\mathcal{N}$ of all nurses working in the CRU. In order to achieve the desired targets on the overutilization and TFAV metrics, CAPTAIN has the ability to provide a Nurse Schedule Plan (NSPlan) that (i) plans for each day d , the number of hours $K_n(d)$ each nurse n should work on day d given the clinic's constraints on shift assignments, and (ii) coordinates participants' needs (skills required on each visit) with nurses' skill sets and availability. Feature (i) permits the model to generate a nursing shift assignment plan that takes into account the possible shift lengths of a CRU (for examples, see Case Studies A and B in Sec. 4.3) as well as the CRU's rules on shift assignments (see the test suite in Sec. 4.1). In the MIP formulation in Appendix F, we omit the nurse staff workload planning constraints, since these must be coded into the MIP in a manner unique to the case study (i.e., rules on shift assignments and shift lengths).

Conceptually, the idea is to ensure that

$\{(K_n(1), \dots, K_n(D + L^{max} + \max_k TFAV_k^{max})) : n \in \mathcal{N}\} \in \Delta$, with the set Δ assuring that all staffing rules are obeyed. The modeling of skill sets allows for the consideration of training/cross-training in addition to hiring/expansion. In (ii), we need to compute decision variables $x_{s,n}(d)$ that determine the number of hours nurse n should be assigned to participants that require skill s on day d of the planning horizon while respecting the allowable shift lengths. Continuity of care is not a concern for the multiple visits of a trial; hence, we do not need to assign the same nurse to all the visits of a participant. Again note that Theorem 3.3 determines the mean workload induced on the nursing skills, but not the nurses themselves (because protocols do not specify which nurse needs to be present for a visit, but rather the skills).

The decision variables $x_{s,n}(d)$ need to ensure that the offered workload on any skill on any day d (expressed in Theorem 3.3) is only assigned to a nurse n that can perform that skill s , where $\mathcal{N}(s)$ denotes the set of nurses that possess skill s . Observe that such a decision not only affects the workload induced on the skill required by this visit, but also on all other skills that this nurse possesses due to shared capacity. The next equation assures that the set of nurses that possess skill s will serve the total anticipated workload for that skill, and that no nurse n will be assigned workload not in their skill set:

$$\bar{W}_s(d) = \sum_{n \in \mathcal{N}(s)} x_{s,n}(d), \forall s \in \mathcal{S}, \forall d \in \{1, 2, \dots, D + L^{max} + \max_k \overline{TFAV}_k^{max}\}.$$

Now that the workload induced on a skill set has appropriately been distributed among the right set of nurses, we can compute the mean daily workload induced on each nurse:

$$\bar{W}(n, d) = \sum_{s \in \mathcal{S}} x_{s,n}(d), \forall n \in \mathcal{N}, \forall d \in \left\{1, 2, \dots, D + L^{\max} + \max_k \overline{TFAV}_k^{\max}\right\}.$$

We allow the system administrator to impose a constraint on the maximum allowable, planned mean overutilization for nurse n on day d , denoted $\bar{O}_n(d)$:

$$\bar{W}(n, d) \leq K_n(d) + \bar{O}_n(d), \forall n \in \mathcal{N}, \forall d \in \left\{1, 2, \dots, D + L^{\max} + \max_k \overline{TFAV}_k^{\max}\right\}.$$

For room resources (e.g., bed, chair, procedure room, exercise room, DEXA scan, etc.), the model is a degenerate case of nurse resources, because there is no need for “room flexibility” or a room “staffing plan.” We model a fixed daily available capacity (of 12 hours in MCRU’s case). From Theorem 3.3, the expression for the mean overutilization hours induced each day on any room resource is easily formulated (see Eq. (51) in Appendix F).

4 Model Validation and Insights

Since the formulations in the previous section rely on some approximations, we first validated these approximations. We validated the BRPlan (i.e., how many people get scheduled on average every day by type), the TFAV metric, and the workload formulation. That is, we validated each part of the math formulations that introduced some assumptions to make the model linear. The validation results are reported in Appendix G.

For both validation and our case studies we used a realistic test suite. Sec. 4.1 gives the test suite. Sec. 4.2 tests the portfolio selection feature of our problem, while Sec. 4.3 tests the improvements made by the BRPlan and NSPlan (which will then be optimized) over the current practice FAS policy. In Sec. 4.3, which considers a fixed portfolio rather than portfolio selection, our objective is to minimize the overall average mean TFAV per patient so that access is as rapid as possible in addition to achieving the required constraints on access delay. For all of the following numerical results (validation and sensitivity/tradeoff analyses), we conservatively employ the $m = 1$ approximation, which highlights a worst case scenario for the approximation. Specifically, in the case where $m = 1$, we lose accuracy in computing the mean number of enrolled participants that will be booked on a given day, because we consider the number of enrollments on the current day, and aggregate all the enrollment information of previous days. Accuracy commonly improves with a larger m .

4.1 Test suite

Our objective is to perform a performance evaluation that is broader than a case study of a specific CRU. Drawing from protocol characteristics from MCRU that are believed to be representative of many sites, we constructed a randomized test suite of protocols, CRU features, and CRU constraints on key metrics, that have characteristics spanning those in practice. We randomized (with uniform distributions) the protocol characteristics of all trials. The ranges on the uniformly distributed parameters of our randomized test suite are provided in Table 1. For our analyses we considered a set of 16 active trials, and for

portfolio selection in Sec. 4.2 we considered an additional 8 candidate trials. This randomization resulted in an instance of 24 trials that varied a lot based on number of visits, treatment durations, and visit window sizes, which implies various trial characteristics and densities. There were also rules on shift scheduling: it is required that a given nurse (1) cannot have two 12 hour shifts (the longest possible shift) two days in a row, (2) cannot be on call (a nonzero capacity) more than three days in a row, and (3) is assigned a total planned capacity of 40 hours in a week. In practice, each performance site may have its own set of rules/preferences, and our MIP methods are flexible enough to incorporate additional requirements.

Note that in practice, candidate trials proposed to the CRU have a fully defined protocol, so the protocol parameters needed for the model will be known prior to the activation of the trial (including the desired number of participants). Estimating enrollments for candidate trials is challenging, but our experience and that of MCRU staff indicates that it can help to use historical enrollment data from trials with similar features. The nurse skill sets were randomized. For each of the 6 nurses, there are 10 possible skills. A probability of 0.5 was used to determine whether or not nurse n has skill s , with i.i.d. samples.

Having formulated our performance metrics linearly in our decision variables, we can constrain them based on managerial inputs. We consider the following constraints: (i) $\bar{O}_r(d)$, the maximum allowable number of overutilization hours for resource r (for nurses and/or rooms) and each day d , (ii) $\overline{TFAV}_k^{avg}$, the upper limit of the population average TFAV for each trial, (iii) $TFAV_{k,t}^{except}(\omega)$, the TFAV exception limiting the percentage of type k participants enrolling on day t that will exceed ω days TFAV, and (iv) $\overline{TFAV}_k^{max}$, the population maximum TFAV for each trial k . The optimization functionality of CAPTAIN was implemented using Visual Studio 2010 with an integrated IBM CPLEX solver. The computer used was an Intel Xeon CPU with 8GB of RAM. For a 60 business day (12 weeks) planning horizon, this program was solved up to a 1% optimality gap in 1 to 6 hours depending on the protocol complexities. As the planning horizon gets larger, the computing time increases exponentially. However, 12 weeks is already considered a very long planning horizon for our application, since CRUs tend to update their portfolios more frequently.

4.2 Tradeoff analysis and portfolio selection

We studied the optimal objective values and the associated portfolio selections under various constraints set on daily overutilization and on average TFAV for the candidate trials. For our planning horizon, we considered 8 candidate trials that could be added to the current portfolio of 16 trials. The SVG of each candidate was set to 1, which emphasizes the operational characteristics. The candidate trials also have randomly generated parameters according to Table 1. The goal of this section is (i) to provide high level validation that the optimization's objective behaves as expected when the constraints on TFAV and overutilization are altered and (ii) to show that the selection of a portfolio is not a simple, intuitive process (because of the complex interactions between protocols). This motivates the critical need for a decision support system to manage a CRU.

A high level validation of our method is seen in the left part of Fig. 5. A CRU can extract greater scientific value from an optimized portfolio when the bound set on the population average TFAV for all candidate trials is increased by the CRU. To illustrate the tradeoff more clearly, in this analysis we treated all 24 trials as candidate trials, allowing any subset to be selected. By having the ability to delay the access of less delay-sensitive participants, the method is able to integrate more trials in our planning horizon's updated portfolio. Moreover, as seen on the left for a fixed population average TFAV for all candidate trials, the four experiments that increase the overutilization bound show that the CRU can extract more scientific value from the portfolio chosen. For example, by allowing more overutilization hours/expenses, high density trials that were difficult to schedule because of the protocol requirements (e.g., those with very long service times, short time windows for subsequent visits to happen, etc.) may become feasible to offer.

The right part of Fig. 5 displays the optimal portfolio selections from the 8 candidate trials (added to the 16 fixed active trials), $\{C1, \dots, C8\}$, each having given bounds on daily overutilization and TFAV. This shows that this decision is neither trivial nor intuitive. By increasing the population average TFAV bound for all candidate trials as indicated in the rows (respectively the overutilization bounds per day, identified for each table), we see that the optimal portfolio selection is not always just monotonically adding new trials to the optimal portfolio selection with lower TFAV bounds (resp., overutilization bounds). For example, with an overutilization bound of 4 hours per day, we would only choose candidate trial C2 when the TFAV bound is 5 days. But when the TFAV bound is increased to 15 days, trial C2 is replaced by C1, C3, and C4 (and we see similar complex behavior when the constraint on overutilization varies). This motivates the utility of the CAPTAIN methodology, since the managerial staff is not able to accurately predict how a given portfolio would impact the site's performance metrics.

4.3 Improvements Relative to First Available Slot (FAS)

To assess the benefits of CAPTAIN relative to the current practice FAS scheduling, we used simulation (for more details see Appendix G). We assumed that no trials would be added at the beginning of our planning horizon (no portfolio selection), because this allows us to isolate the difference between an optimized BRPlan and FAS scheduling for a given portfolio generated by the randomized test suite (Table 1). To illustrate how using CAPTAIN as a booking/planning tool to schedule participants is itself very helpful to control and improve the CRU metrics, we minimized the average mean TFAV across all participants over all days of our planning horizon, letting the solver find an optimized BRPlan that satisfies the constraints set on mean overutilization, population mean TFAV, TFAV-exception, and population maximum TFAV for each trial. This involves multiplying the TFAV of each trial by its mean arrival rate and using Remark 3.2 to write the objective function. Minimizing the average total TFAV across all participants achieves as rapid access as is possible on average. Note that the objective could also be to minimize the aggregated workload planned for service on each day, which encourages high throughput, but to be meaningful this also requires the bound on overutilization (see constraint (i) listed in Sec. 4.1).

Fig. 6 presents a visual illustration of the capabilities and potential of CAPTAIN for one representative problem instance (portfolio of trials) from our randomized test suite in Table 1. The upper left of Fig. 6 shows the population average TFAV, the population standard deviation TFAV, and the population maximum TFAV for each trial in a CRU that operates under the current FAS daily level booking policy with a total overutilization limit of 4 hours per day (over all human resources). This was illustrated in the example of Fig. 2 (left), but we added the population maximum TFAV for each trial on the graph. In this problem instance, the FAS policy resulted in a total of 1.4 overutilization hours per day on average over all resources (given an overutilization limit of 4 hours per day over all resources). The capability of CAPTAIN allows the user to input any desired target on mean overutilization (per resource and per day), but in order to show in a fair way the improvements made by CAPTAIN, we kept the mean overutilization to the same level as the FAS policy in Fig. 6. The upper right of Fig. 6 shows by protocol the population average TFAV, the population standard deviation TFAV (from simulation after the BRPlan has been optimized), and the population maximum TFAV for each trial in a CRU that operates under CAPTAIN's daily level BRPlan methodology and achieves on average the same overutilization as the FAS policy while also meeting certain targets on population average TFAV for each trial. The horizontal dashed lines show the population average TFAV targets we set on each trial in our optimization. Although we are able to achieve such targets (which gives us the control to prioritize needier participants and/or high-valued trials), we realize that the population standard deviation TFAV for trials with "poor" access (e.g., trials 8-13) tends to be larger than those under FAS policy. In contrast, trials with "good" access (e.g., trials 1-5) will have lower TFAV variability than under the FAS policy. The explanation is that trials with higher population average TFAV targets will have a larger number of successive days with no BRPlan reservations (since they are not prioritized) which significantly increases TFAV variability between participants. Decision makers may want to avoid such scenarios. For example, it is likely that guaranteeing 25 days access on average to their participants will result in a high number of them waiting longer than that due to variability.

One will often want to reduce the TFAV-exception rate for each trial to guarantee that no more than a certain percentage of type k participants will exceed a given target. We added those constraints to the optimization (while keeping the previous overutilization and population average TFAV constraints) by selecting a TFAV-exception target of $\left[\overline{TFAV}_k^{avg} + \sqrt{TFAV_k^{avg}} \right]$ and a TFAV-exception rate of 10% for each trial k . The results are illustrated in the lower left of Fig. 6. Although we see a slight increase in the population standard deviation of trials with high access, we were able to decrease the access variability for participants in trials that have a larger average TFAV target. However, the population maximum TFAV of those trials has increased compared to the population maximum TFAV of the upper right of Fig. 6, even though we have decreased the population standard deviation of those trials. Notice that by definition, the TFAV-exception controls the percentage (or number) of participants exceeding a TFAV target, but not the number of days that those participants will exceed the target. Hence, in the lower left of Fig. 6, we may have a small number of participants from trials 8-13 that exceed the $\left[\overline{TFAV}_k^{avg} + \sqrt{TFAV_k^{avg}} \right]$ target. However the few that do are likely to exceed it by a lot, thereby increasing in the

population maximum TFAV. Again, decision makers may want to avoid such situations and also be able to control the population maximum TFAV such that there isn't even a single patient that is predicted to wait extremely long for first visit.

In the lower right of Fig. 6, we see the portfolio TFAV results of a BRPlan that assures (i) population average TFAV targets per trial, (ii) TFAV-exception targets for each trial, and (iii) population maximum TFAV targets per trial (indicated by solid horizontal lines). Of course, reducing the population maximum TFAV while keeping the TFAV-exception % the same will result in a decrease in the population standard deviation.

We observe significant benefits from using the BRPlan booking methodology rather than the FAS policy. With a mean overutilization limit of 1.4 hours, the BRPlan was able to reduce the population maximum TFAV (the longest time any participant will have to wait) from over 55 days (using the FAS policy, upper left of Fig. 6) down to 40 days (lower right of Fig. 6). Also, we see that we were able to limit the population average TFAV for each trial to the desired limit, therefore significantly increasing access to care for more urgent participants. By adding constraints on the TFAV-exception rate, we are able not only to reduce TFAV variability between participants of the same trials, but we are also able to give the managerial staff some valuable information: they will then be able to advertise that a given trial guarantees $y\%$ (e.g., 90 %) of its participants will have access within x days (e.g., 15 business days) after enrollment. Having this information may increase enrollment (by increasing the perceived value/service), increase participant satisfaction, and decrease overall trial duration. We notice that under both scheduling methodologies (FAS and BRPlan), we have the same amount of daily capacity at our disposition, but we are able to limit average waiting times to significantly shorter amounts while also reducing access variability for each trial. This happens because the BRPlan is able to reduce the workload predictable and unpredictable variability by packing participant visits more efficiently while the FAS rule generates big "holes" (and workload variability) in the schedule. The BRPlan levels the daily workload (and reduces its variability) while also providing overall better access, which results in an increase in resource utilization for a transient planning horizon (see also the FAS and CAPTAIN-Base Case rows in Table 2).

Taking this one step further, we now allow the MIP to also optimize the NSPlan. The number of hours that each nurse will work on each day will now be optimized (under the same parameters and constraints in the lower right part of Fig. 6) while obeying a certain set of rules (defined generally as in Sec. 3.5). In the following case studies, we use the same set of rules reported in the description of our test suite (Sec. 4.1). In Case Study A, we consider three types of shifts for an on-call nurse: 4, 8, or 12 hour shifts. In Case Study B, we have more shift options/flexibility: 4, 6, 8, 10, 12 hour shifts. Additionally, we consider shifts of length zero in our constraints for days when the nurse would have a day off. Our FAS case study and CAPTAIN Base Case consider a fixed (suboptimal) NSPlan (the same one used in our calculations in the upper left of Fig. 6 for FAS, and lower right of Fig. 6 for the CAPTAIN Base Case). Note again that in all of our 4 cases (rows in Table 2), we are considering the same volume of arrivals and the same nurse total capacity over all days of our transient planning horizon. We see that the largest marginal improvement from the FAS policy (in terms of the average and standard deviation of the utilization and TFAV metrics) is

the optimization of a BRPlan while keeping the nurse schedules the same (see the first two rows of Table 2).

But we can provide even more improvements by also creating an NSPlan (that obeys the same set of nurse scheduling rules) that is more flexible and appropriate for the optimized BRPlan. Our forecasting methodologies create the relation between fulfilled demand and offered workloads on skills, and the NSPlan links the right amount of nurse capacity based on the skill sets needed on each day. Although it creates only small improvements on our metrics, it is worth noting that having more shift options (therefore flexibility in scheduling nurse capacity with our NSPlan) provides an increase in utilization (and workload smoothing), while increasing access for patients (see Case Studies A and B in Table 2). In equilibrium, higher utilization generates longer waiting times for a fixed throughput; however, in our transient case, we are able to increase throughput (throughout our planning horizon) by providing shorter waiting times which translates into higher utilization within our transient planning horizon.

We have explained the use of CAPTAIN for CRUs which are faced with the decision of which portfolio mix of candidate trials should be performed (in addition to the current portfolio), given a fixed infrastructure and set of nurses with specific skills. Because we validated the ability of CAPTAIN to accurately optimize the transition from a prior set of appointments on the books (i.e., under FAS), this also assures us that CAPTAIN can handle transitions caused by future periods (see Appendix C for further details). This decision support tool could also be used for physical capacity expansions and hiring purposes. It can answer the question: to what level of capacity should we expand our resources in order to meet our objectives for scientific value, service/access levels for participants, and staff overutilization? In other words, given any portfolio of trials and service level constraints, CAPTAIN could determine the minimum staff and room capacities required to conduct this portfolio.

5 Conclusion

In contrast to current practice in CRUs, which is often based on common sense and experience, this new planning model, CAPTAIN, increases operational efficiency (and thus cost effectiveness), increases the ability to extract scientific value by optimizing the selection of the clinical trials performed, and provides a high level of access with limited waiting times that are planned to match the participants' needs. This planning system coordinates care resources and participant visits via novel forecasting and optimization algorithms. This paper provides a proof of concept for a general decision support tool that can enable a CRU to gain control over key performance metrics. Based on a rolling planning horizon, CAPTAIN (i) determines the optimal mix of protocols to perform, (ii) forecasts and controls the workloads that will be placed on resources (staff and rooms), (iii) forecasts and controls overutilization in part by creating efficient nurse schedules and participant/staff coordination, (iv) provides a "system optimal" daily level scheduling plan via the Booking Reservation Plan (BRPlan), (v) optimizes the allocation of resource capacity, (vi) forecasts and controls the participants' waiting times to first visit, and (vii) answers "what if" questions by allowing CRUs to enter desired constraints on specific metrics to see the impact

on others. As a non-stationary model, CAPTAIN can begin with a base case representing the current conditions and all the existing appointments made. From this, it can then select the potential clinical research trials that should be conducted without generating excessively long access delays (to first visit) and high overutilization costs. To guide appointment scheduling, CAPTAIN generates a BRPlan that is time-varying, maintaining effective operations from one planning period to the next.

Whereas this preliminary work targets strategic planning, future research can focus on appointment scheduling at the time of day level. Therefore, as future research, the methodology will benefit from another layer of modeling and resource allocation to bridge from the planning level to the execution level to schedule time of day appointments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgement

Partial funding for this work has been received from NSF Grants CMMI 1233095, 1548201 and NIH CTSA grant UL1TR000433. The funding organizations had no role in the design or conduct of this research. Dr. Deglise-Hawkinson acknowledges partial support from a Seth Bonder Fellowship.

Biographies

Jivan Deglise Hawkinson is an OR senior Analyst in the Revenue Management (RM) department at American Airlines (AA). His interests include data driven forecasting methodologies and leveraging them to create optimized decision support tools. His current work is focused on developing and improving the forecasting tools of AA, to make them more dynamic, while using machine learning methodologies. He was Chief Scientist at Lean Care Solutions (2015-2018), a healthcare startup company focused on creating decision support tools for hospital readmissions prediction and prevention. He received his PhD from the University of Michigan Industrial and Operations Engineering department (2010-2015) focusing on admission control and large stochastic queuing networks applied to various healthcare systems.

David Kaufman is an Assistant Professor at the University of Michigan-Dearborn, College of Business, where he teaches courses in Decision Sciences. He holds a Ph.D. in Industrial and Operations Engineering from the University of Michigan. Prior to joining the faculty at Dearborn, Dr. Kaufman was a lecturer at the University of Michigan, where he taught courses in operations research, corporate finance, and financial engineering. His previous teaching experience also includes a course in stochastic processes at Cornell University. Dr. Kaufman's previous industry experience includes product development for RiskMetrics Group, a financial risk management company that was acquired by MSCI. His research interests are in stochastic processes and decision models for systems where variability and uncertainty play an important role in design, analysis, and management.

Dr. Blake Roessler is Professor Emeritus of Internal Medicine University of Michigan Hospitals-Michigan Medicine. He was also the director of the Michigan Clinical Research

Unit (MCRU) at the time of this research. He has developed an ongoing interest in operational engineering of clinical research systems as well as clinical research applications of cell and gene therapy. As a rheumatologist with specialties in gout and uric acid metabolism, he has been affiliated with multiple hospitals, including Michigan Medicine and the Veterans Affairs Ann Arbor Healthcare System. He received his medical degree from University of Cincinnati College of Medicine and practiced medicine for more than 20 years before his retirement a few years ago.

Mark Van Oyen is a Professor of Industrial and Operations Engr. (IOE) at the Univ. of Michigan. His interests include the analysis, design, prediction and control of stochastic systems. His current research emphasizes stochastic systems, optimization, and prescriptive analytics for healthcare operations and medical decision-making. He co-authored papers that won numerous awards. He has served as Associate Editor for Operations Research, Management Science, Naval Research Logistics, and IIE Transactions, and IIE Trans. Healthcare Syst. Engr. and Senior Editor for Flexible Services & Manufacturing. He was a faculty member of the Northwestern Univ. Sch. of Engr. (1993-2005) and Loyola Univ. of Chicago's Sch. of Bus. Admin. (1999-2005). He has received grant funding from the NSF, ONR, NIH, EPRI, ALCOA, General Motors, and the VA. He received his Ph.D. from Electrical Engr. Systems from the Univ. of Michigan and has worked for GE Corporate R&D.

References

- Ahmadi-Javid A, Jalali Z, Klassen KJ. 2017 Outpatient appointment systems in healthcare: A review of optimization studies. *Eur. J. Oper. Res* 258(1) 3–34.
- Ahuja V, Birge JR. 2016 Response-adaptive designs for clinical trials: Simultaneous learning from multiple patients. *Eur. J. Oper. Res* 248(2) 619–633.
- Akkan C 1997 Finite-capacity scheduling-based planning for revenue-based capacity management. *Eur. J. Oper. Res* 100(1) 170–179.
- Anderson JR, Matessa M. 1997 An overview of the epic architecture for cognition and performance with application to human-computer interaction. *Human-Computer Interaction* 12(4) 391–438.
- Asmundsson J, Rardin RL, Turkseven CH, Uzsoy R. 2009 Production planning with resources subject to congestion. *Nav. Res. Log* 56(2) 142–157.
- Berry DA 1978 Modified two-armed bandit strategies for certain clinical trials. *J. Am. Stat. Assoc* 73(362) 339–345.
- Cayirli T, Veral E. 2003 Outpatient scheduling in health care: A review of literature. *Prod. Oper. Manag* 12(4) 519–549.
- Condotta A, Shakhlevich NV. 2014 Scheduling patient appointments via multilevel template: A case study in chemotherapy. *Oper. Res. for Health Care* 3(3) 129–144.
- Deglise-Favre-Hawkinson J 2015 Access and resource management for clinical care and clinical research in multi-class stochastic queueing networks. Ph.D. thesis, University of Michigan.
- Deglise-Hawkinson J, Helm JE, Huschka T, Kaufman DL, Van Oyen MP. 2018 A capacity allocation planning model for integrated care and access management. *Prod. Oper. Manag* 27(12) 2270–2290. [PubMed: 30930608]
- Diamant A, Milner J, Queresby F. 2018 Dynamic patient scheduling for multi-appointment health care programs. *Prod. Oper. Manag* 27(1) 58–79.
- Farrell B, Kenyon S, Shakur H. 2010 Managing clinical trials. *Trials* 11(1) 78. [PubMed: 20626885]
- Felli JC, Anderson WH, Kremidas JP, Ruberg SJ. 2007 A semi-Markov model for patient progression through clinical trials. *Eur. J. Oper. Res* 176(1) 542–549.

- Gocgun Y, Gbate A. 2012 Lagrangian relaxation and constraint generation for allocation and advanced scheduling. *Comput. Oper. Res* 39(10) 2323–2336.
- Gocgun Y, Puterman ML. 2014 Dynamic scheduling with due dates and time windows: An application to chemotherapy patient appointment booking. *Health Care Manag. Sc* 17(1) 60–76. [PubMed: 24114392]
- Grunow M, Günther HO, Yang G. 2004 Development of a decision support model for scheduling clinical studies and assigning medical personnel. *Health Care Manag. Sc* 7(4) 305–317. [PubMed: 15717815]
- Gupta D, Denton B. 2008 Appointment scheduling in health care: Challenges and opportunities. *I.I.E. Transactions* 40(9) 800–819.
- Gupta D, Wang L. 2008 Revenue management for a primary-care clinic in the presence of patient choice. *Oper. Res* 56(3) 576–592.
- Heine C, Herrler R, Petsch M, Anhalt C. 2003 Adapt: adaptive multi-agent process planning & coordination of clinical trials. *Proceedings of the 2003 Americas Conference on Information Systems* Citeseer, 1823–1834.
- Helm JE, Van Oyen MP. 2014 Design and optimization methods for hospital admissions. *Oper. Res* 62(6) 1265–1282.
- Hsu H-M, Wang W-P. 2001 Possibilistic programming in production planning of assemble-to-order environments. *Fuzzy sets and Systems* 119(1) 59–70.
- Hulshof PJH, Boucherie RJ, Hans EW, Hurink JL. 2013 Tactical resource allocation and elective patient admission planning in care processes. *Health Care Mgt. Sci* 16(12) 152–166.
- Johansen MA, Mayer DK, Hoover HC Jr. 1991 Obstacles to implementing cancer clinical trials. *Seminars in Oncology Nursing* 7 (4) 260–267. [PubMed: 1775776]
- McGarvey BM, Dynes NJ, Lin BC, Anderson WH, Kremidas JP, Felli JC. 2007 A discrete event model of clinical trial enrollment at Eli Lilly and Company. *Simulation Conference, 2007 Winter IEEE*, 1467–1474.
- Meinert CL, Tonascia S. 1986 *Clinical trials: design, conduct and analysis*. Oxford University Press, USA.
- Mula J, Poler R, Garcia-Sabater JP, Lario FC. 2006 Models for production planning under uncertainty: A review. *Int. J. Prod. Econ* 103(1) 271–285.
- Patrick J, Puterman ML, Queyranne M. 2008 Dynamic multipriority patient scheduling for a diagnostic resource. *Oper. Res* 56(6) 1507–1525.
- Phillips C, Stein C, Wein J. 1995 Scheduling jobs that arrive over time Akl SG, Dehne F, Sack JR, Santoro N, eds., *Algorithms and Data Structures. WADS 1995. Lect. Notes Comput. Sc*, vol. 955 Springer, Berlin, Heidelberg, 86–97.
- Pinedo M 2008 *Scheduling: theory, algorithms, and systems*. Springer Verlag.
- Roland CL, Litka P. 2007 *Clinical trial operations* Evens RP, ed., *Drug and Biological Development*. Springer, Boston, MA, 178–201.
- Rubin ER, Lazar D, Gaich N, Haray D. 2007 The clinical trials landscape: Limitations, strengths, and promise.
- Sauré A, Begun MA, Patrick J. 2019 Dynamic multi-priority, multi-class patient scheduling with stochastic service times. *Eur. J. Oper. Res* to appear, 1–33.
- Sauré A, Patrick J, Tyldesley S, Puterman ML 2012 Dynamic multi-appointment patient scheduling for radiation therapy. *Eur. J. Oper. Res* 223(2) 573–584.
- Talluri KT, Van Ryzin GJ. 2006 *The theory and practice of revenue management*, vol. 68 Springer Science & Business Media.
- Turkan A, Zeng B, Lawley M. 2012 Chemotherapy operations planning and scheduling. *IIE Trans, on Healthcare Sys. Eng* 2(1) 31–49.
- Varma VA, Uzsoy R, Pekny J, Blau G. 2007 Lagrangian heuristics for scheduling new product development projects in the pharmaceutical industry. *J. Heuristics* 13(5) 403–433.

Fig. 1.
Example of the current practice First Available Slot (FAS) booking methodology.

Fig. 2.

Impact of a trial's protocol density on it's own TFAV as well as the rest of the portfolio of trials. Left: Population average and standard deviation of the TFAV under the FAS policy for a given instance of the randomized test suite in Table 1. Right: TFAV mean and standard deviation of trial 3 and the same portfolio instance averaged across all 15 other protocols as the mean workload per visit of trial 3 varies.

Fig. 3.

BRPlan scheduling example with daily enrollments $A_{k,t}^{(w)}$ realizations and daily capacity $\theta_{k,d}$.

Fig. 4. Visualization of the enrollment history taken into account when we condition the expectation of the number of patients that enrolled on day 11 and are scheduled on a future day d under different values of the m parameter.

Fig. 5.

Left: Tradeoff analyses of the Scientific Value Gained vs. the bound set on population mean TFAV (in days) for 24 candidate trials and the overutilization bound in hours per day over all nurses; Right: Optimal portfolio selection with respect to population mean TFAV for 8 candidate trials when the overutilization bound is 2, 4, and 6 hours per day over all nurses.

Fig. 6.

Simulation and optimization analysis of CAPTAIN's improvements and controls over TFAV variability compared to current practice under the same overutilization level of 1.4 hours per day over all resources. Upper left - FAS scheduling; Upper right - BRPlan with average TFAV constraints indicated by dashed lines; Lower left - BRPlan with constraints of upper right graph and a TFAV-exception of 10%; Lower right - BRPlan with constraints of lower left graph and a TFAV maximum as indicated by solid lines.

Table 1

Test suite defined by trial protocol features (upper part of the table), CRU features (middle part of the table), and targets/bounds set on the main metrics (lower part of the table).

Parameter	Range
Workload induced on a resource for a given visit (in hours)	1 - 6
Time between successive visits (in business days)	1 - 15
Visit time windows (in $\pm$ days)	0 - 10
Number of visits	1 - 10
Enrollment rates (in participants/day)	0.1 - 1
Number of active trials	6 - 16
Number of candidate trials to consider accepting	0 - 8
Number of nurses	6
Number of nurse skill sets	10
Possible shift lengths (in hours/day)	0, 4, 8, or 12
Number of room types	6
Room capacity (in number of hours per day for each type)	12 - 48
Daily overutilization target over all resources (in hours)	0 - 6
Population Average TFAV target $TFAV_k^{avg}$ (in business days)	10 - 25
Mean percentage of participants exceeding a given TFAV target (in %)	0 - 50
Population Maximum TFAV target (in business days)	$TFAV_k^{avg} - 40$

Table 2

Calculations of (i) the mean total resource utilization (averaged over all days of our planning horizon), (ii) the standard deviation of the total resource utilization (averaged over all days of our planning horizon), (iii) the mean TFAV in days averaged over all 16 active trials, (iv) the standard deviation of the TFAV averaged over all 16 active trials, and (v) the maximum TFAV averaged over all 16 active trials, for the following cases: (a) the FAS policy, (b) CAPTAIN's Base Case (BRPlan optimization with a fixed NSPlan), (c) CAPTAIN'S Case Study A (BRPlan and NSPlan optimization with 0, 4, 8, and 12 hour shift options), and (d) CAPTAIN's Case Study B (BRPlan and NSPlan optimization with 0, 4, 6, 8, 10, and 12 hour shift options). Note that the constraints specified by the lower right part of Fig. 6 were enforced for the CAPTAIN results.

		Utilization (%)		TFAV (days)		
		Average	Standard deviation	Average	Standard deviation	Maximum
FAS		79.3%	12.6%	24.5	4.1	56.2
CAPTAIN	Base Case	86.1%	8.2%	14.2	3.6	38.6
	Case Study A	88.9%	7.4%	13.0	3.1	32.8
	Case Study B	90.1%	7.1%	12.8	2.9	29.9