Article

A Robust Predictive Resource Planning under Demand Uncertainty to Improve Waiting Times in Outpatient Clinics

Journal of Health Management
19(4) 563–583

SAGE Publications
sagepub.in/home.nav
DOI: 10.1177/0972063417727627
http://jhm.sagepub.com

Jyoti R. Munavalli[1]
Shyam Vasudeva Rao[2]
Aravind Srinivasan[3]
Usha Manjunath[4]
G. G. van Merode[5]

Abstract

Background and context: Resource planning is performed ahead of time within outpatient clinics (OPC). Due to local control of operations (department-centric decision-making) and limited resources, OPCs cannot handle high variability and uncertainty in demand. There is always a difference between planning and reality, and this leads to operational problems such as excessive waiting times. The OPCs often react to the situation when problems are encountered and reaction times play an important role in determining patient waiting times.

Objectives: To propose a predictive resource planning that incorporates variability in the short term with the OPC-wide perspective, not department-centric.

Methodology: The process and patient data were collected from the OPC under study by observation, interviews and from the records of the hospital management information system. A resource planning model (RPM) was developed that matched resources according to demand in short term. A mathematical model with outputs resource plan for a day was formulated utilizing Takt time (the average time a patient needs to move out of the OPC system) management that is used in Toyota Production System (TPS), to allocate resources to all the departments. Using a Discrete Event Simulation Model,

[1] Research Associate, Maastricht University Medical Centre, Maastricht, The Netherlands.
[2] Cofounder, President and CTO, Forus Health, Bangalore, Technical Director, Maastricht University Medical Centre, Maastricht, The Netherlands.
[3] Administrator, Aravind Eye Care, Madurai, Tamil Nadu, India.
[4] Professor, Institute of Health Management Research, Bangalore, Karnataka, India.
[5] Professor of Logistics and Operations Management in Health Care, Maastricht University Medical Centre, Maastricht, The Netherlands.

Corresponding author:
Jyoti R. Munavalli, Research Associate, Maastricht University Medical Centre, Maastricht, The Netherlands.
E-mail: jyothimunavalli@gmail.com

the effects of predictive resource planning with different reaction times on waiting times and cycle times were analyzed. The resource plans were implemented in the OPC of Aravind Eye Hospital, Madurai, Tamil Nadu, India, that has high patient volumes and random patient arrivals.

Results and discussion: The simulation and implementation results indicate that predictive resource planning is robust and improves waiting times, and cycle times in OPCs. Study findings confirm that the predictive planning model reduces the average waiting time by 43.4 per cent during simulation and by 41.1 per cent during its implementation. The reduction in standard deviations in waiting times indicate reduction of unregulated waiting times. The OPC scheduled 28 resources throughout the day, whereas with predictive resource planning, the number of resources varied between a minimum of 12 to a maximum around 30–34 resources.

Conclusions: The OPCs currently match demands to their supply, while matching resources to varying demand in short term; throughout the OPC (all departments) improves patient flow, and minimizes waiting time and cycle time. Previously, Takt time management (TTM) has applied to systems with even and stable demand; in this study, it has been applied to stochastic demand.

Implications: This planning model helps the management to identify resource requirements: types of resources and number of resources, for the future demand growth and expansion. It can probably be extended to general hospitals by considering their demand forecast, precedence constraints and workflow complexities.

Keywords

Predictive resource planning, waiting time, Takt time management, reaction time, outpatient clinic, demand-supply

Introduction

Resource planning and control are becoming extremely important in outpatient clinics (OPCs). Due to expanding patient demand, greater patient expectations and increasingly complex patient flow, OPC systems are under constant pressure to provide quality care despite limited resources (Huang, 1994; Pillay et al., 2011; Zhu, Heng & Teow, 2012). Patient waiting time is the major reason for complaints and patient dissatisfaction and plays a crucial role in quality management.

The OPCs face operational problems because of the variability and uncertainty in demand and service times, and often resource planning are based on aspects such as (i) average patient demand, (ii) the resource scheduling performed ahead of time and (iii) local optimization. In most clinics, resources in OPCs are planned and managed through a simple deterministic approach using average demand and average service times (Harper, 2002). Patient arrivals are not uniformly distributed but mostly Poisson (with patient demand changing from hour to hour) (Hassan, Zaqhloul & Mokhtar, 2005). Often OPCs fail to incorporate or overlook this variability and uncertainty in short terms (throughout the day), during planning (Nguyen, Sivakumar & Graves, 2015). As a result, OPCs face inefficiencies, long patient waiting times and cycle times, and resource under-utilization, which in turn affects patient satisfaction adversely (Vermeulen et al., 2009).

The number of resources required in an OPC is determined at an aggregate level planning (Vissers, Bertrand & de Vries, 2001). The OPCs forecast demand based on experience, historical data or sometimes advanced analytics. Low demand clinics plan resources up to bimonthly, whereas high demand clinics plan monthly or weekly (Mansdorf, 1975; Yurko et al., 2001). Planning depends on long-term forecasts, whereas variability in demand occurs in short term. The OPCs being open loop systems are influenced by its environment and experience variability and uncertainties that are caused by late, early or random patient arrivals, varied service times and unpredictable clinical pathways (Vermeulen et al., 2009). Resource planning based on incomplete demand information does not fully reflect reality. The mismatch between planning and reality results in either long waiting time or under-utilization of resources. Often OPCs view these frustrating delays as a capacity problem, whereas delays are likely caused by poor capacity or resource management (Voort, Merode & Berden, 2010). Seasonal variations are predicted and managed by increasing capacity (e.g., part-time doctors) (Edward et al., 2008; Molema et al., 2007).

The disparate departments in OPCs plan and control their operations at the departmental level (local). Every department controls its own patient flow and resources. Decisions are taken often without coordination with other departments, because of which patients from upstream departments are pushed to downstream departments that are not ready to service them. Thus, patients wait in some departments, and resources remain idle in other departments. Ludwig, Merode and Groot (2010) provide an insight on the relation between departmental efficiency and hospital efficiency. The local optimization in departments might improve departmental efficiency but does not necessarily improve OPC-wide efficiency. The departmentally optimized OPCs match patient demand with their services, whereas an OPC to be patient-centric should match its resources to patient demand. Van Merode Groothuis and Hasman (2004) suggest the use of short-term planning when demand is non-deterministic. Planning and control approaches that are commonly used are inadequate as they are not demand-driven and lack synchronization. Therefore, there is a need to design an OPC system that synchronizes patient flow between the departments and determines the service pace based on actual patient demand with minimum waiting time.

Toyota Production System (TPS) is a world leader in industrial production that controls its waiting times effectively. The TPS applies just-in-time for a system, not for a single department, that is, a department should not work either slower or faster than the other departments in the line. The TPS sets a pace for product flow by applying *Takt time management* (TTM). 'Takt time' is derived from the German word *Taktzeit* for pace or rhythm. It is the desired time between units of output, to be synchronized to the customer demand. The TPS plans, schedules and controls its resources, raw materials, etc., around the required Takt time. The TPS also apply line balancing (dividing workload as evenly as possible) to increase the overall productivity (Alvarez & Antunes Jr., 2001; Day, Dean, Garfinkel & Thompson, 2010; Eswaramoorthi et al., 2012; Sandanayake & Oduoza, 2009). Takt time can be used in a production management system termed as TTM. It fits well in assembly line systems with few product types, known (stable or even) demand, flexible, multi-skilled workforce, single routing and identical work times. A Takt time-based system is demand-driven. Therefore, it eliminates overtime and overproduction and stabilizes the system. The TPS utilizes a combination of push and pull to reach and maintain continuous process flow in order to reduce work-in-process (WIP) (Chan et al., 2014; Hopp & Lovejoy, 2012; Liker, 2004, p. 330).

This article is based on the case study of resource planning at the OPC in Aravind Eye Hospital (AEH). The AEH is a renowned eye care hospital in Madurai, South India, that provides patient-centred care (Brilliant & Brilliant, 2007; Chaudhary, Modi & Reddy, 2012; Mehta & Shenoy, 2011, p. 336). The AEH performed 401,529 surgeries and treated 2,396,864 outpatients during 2014–2015 (Activity-Report, 2014–2015). The hospital runs with assembly line efficiency, strict quality norms, standardization, cost control and above all high patient volume. The resources (ophthalmologists and paramedical staff) are well trained and dedicated. The OPC predicts its patient demand and uses it for decisions on staff costing and recruiting and for creating awareness among managers. The AEH has a resemblance to an assembly line system (a line of workers and equipment along which a product being assembled passes consecutively from operation to operation until completed) (Andersen & Poulfelt, 2014; Chaudhary et al., 2012;

Natchiar, Thulasiraj & Sundaram, 2008; Rangan & Thulasiraj, 2007). The patients in the OPC move through various departments that perform specific and successive tasks. However, some aspects of the OPC differ from assembly line systems like local control of operations in departments and uneven patient demand. The patient arrivals are random (no appointment systems used) and independent. This makes patient demand highly variable and uncertain. Additionally, it also presents no control on input and constraints on output, as the OPC provides care for all the arrived patients on the same day (Cayirli & Veral, 2003; Gupta & Denton, 2008).

The functional and operational structure of the OPC in AEH with respect to resource scheduling is shown in Figure 1. Accordingly, in this study, we consider an OPC with two identical units. Departments such as new registration (NR) and review registration (RR) are common to both units and each unit has five departments: Vision (V), refraction (RF), tension (TN), dilatation (DL) and preliminary and final examination (PE & FE). For clarity, only unit 1 is shown in Figure 1. All the queues (1–7) are first come, first served (FCFS) basis. The patient flow arrows show the possible pathways for new and review patients.

The OPC in Figure 1 runs according to the rules in Box 1. Managers are the controllers (C) who plan, schedule and control the activities of the departments (local). Different managers are responsible for scheduling the resources (r) such as ophthalmologists and paramedical staffs. The OPC schedules the ophthalmologists once a month, considering their availability after academic (teaching and research) activities and surgery schedules. Departments of both units have the same number of resources, which remains fixed throughout the day and month (rule 3). During peak hours, managers (controllers) apply rules 4 and 7 to the situation. Based on their experience, they increase resources in their departments or shorten the lunch breaks of the staff, to control waiting times, $w(t)$. When upstream departments work faster, the patients flood the downstream departments, which are unready to handle the increased workload. Similarly, when upstream departments work slower than downstream departments, the latter starve (wait for patients). The lack of coordination among departments increases the unregulated waiting time in some departments and under-utilization in some departments.

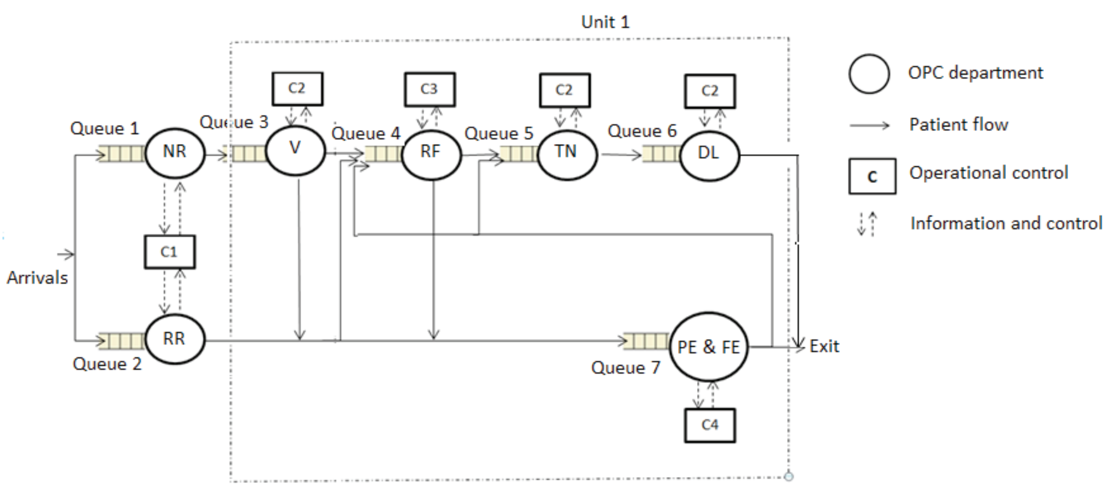

Figure 1. Functional Structure and Operational Control in the OPC System of AEH

Source: Derived from workflow of OPC in AEH.

Box 1. Rules Followed in OPC in AEH

Rule 1: All patients are provided with the service/care the same day of arrival.
Rule 2: All the queues are first come, first served basis.
Rule 3: Resource planning and scheduling time window is one month.
Rule 4: Numbers of resources are fixed throughout day and month.
Rule 5: Resources are scheduled by managers.
Rule 6: Number of resources in unit 1 = number of resources in unit 2
Rule 7: If (queue *n* > threshold workload) → (Resources '*r*' is added to *n*th department D_n such that $r \leq R_T$, the total number of resources) else shorten the lunch times of the already working resource.
Rule 8: Patients should complete RF and PE before TN.
Rule 9: Control of operations is local (department-centric)

The patient workflow in the OPC starts with registration and finishes with the final examination. A patient is in one of these states: waiting state, processing state or finish state. The patient moves through various departments by pathways: NR-V-PE-RF-TN-DL-FE for new patients and RR- PE-RF-TN-DL-FE for review patients. The order of departments RF and PE can be interchanged. Around 5 per cent of the total number of patients exit after the PE. We define five states (in perspective of operations) for a department. S1 → initial state, S2 → waiting state, S3 → regular state, S4 → reactive state and S5 → finish state. Figure 2 shows the transition of a department in different states.

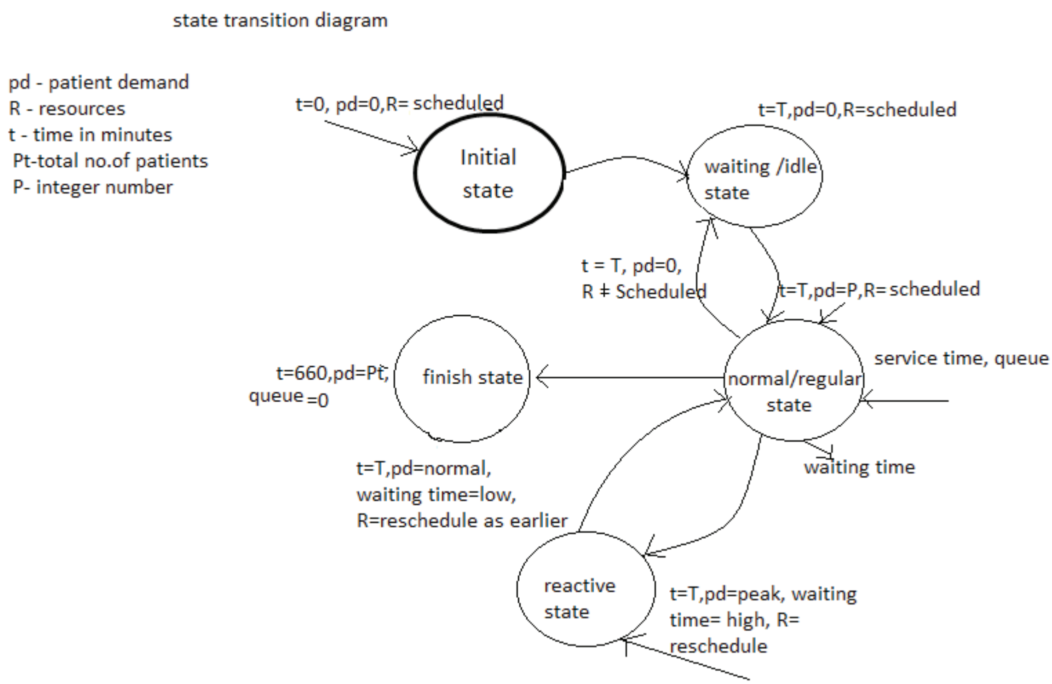

Figure 2. State Transition Diagram for the Department

Source: Derived from departmental workflow of OPC in AEH.

The department is in initial state S1 at the time $t = 0$ (start of the day) with patients not yet arrived. The department moves to waiting state S2 at $t = T$ if either the patient waits for resources or vice versa. The department is in regular state S3 when patients arrive at $t = T$ and are serviced by resources. During peak time, due to high patient demand, the department has the maximum WIP that the OPC can handle. Now the department moves to reactive state S4. The manager who controls the operations in departments responds to the change within limited knowledge (local and available) and follows Rule 7. Once the patient demand becomes regular, the department moves from S4 to regular state S3 and finally, when all the patients are serviced at the end of the day, the department moves to finish state S5.

The time taken by the manager to change from the regular to reactive state (S3 to S4) of the department is the reaction time $r(t)$, and it affects the waiting time. The reaction time varies, as the resources to be transferred to the departments may be busy performing tasks elsewhere in different OPCs. Therefore, the reaction time largely depends on the resource availability at the time of need and the kind of measures taken. The corrective measures improve waiting time in departments but not necessarily the cycle time of patients.

In current practice, though the OPC in AEH is efficient, still it faces operational problems such as long waiting times and cycle times and under-utilization. The OPC, like other hospitals, is an open loop system with uneven demand, lacks synchronization between departments and does not have demand-driven resource planning.

As our main contribution, we present a robust predictive resource planning approach to adapt resources according to demand variability in short-term (hour–hour) and synchronize patient flow between departments through organization-wide perspective in planning. That is, OPC system that services patients based on actual demand (pull system) rather than based on projected or average demand (push system). In this approach, the resources are planned and scheduled in short term (an hour timeslot) based on actual demand. We utilize TTM to implement demand-driven and organization-wide resource planning. Like in fast food chains, raw materials are stored based on projected demand (push), whereas burgers are prepared on customer order (pull). In this approach, we integrate organization-wide planning with almost real time (near to actual demand) planning. Additionally, we present an approach to determine an optimal number of resources throughout a day. By setting a pace between demand and service, waiting times and cycle times are minimized. We extensively evaluated our predictive resource planning in a precisely simulated environment. The planning model is solved using integer linear programming (ILP) to identify the required number of resources. We evaluated simulation under various scenarios. Later, we implemented the resource plans obtained from planning model in the OPC of AEH.

The remainder of this article is organized as follows: In the second section, a robust predictive RPM that incorporates short-term demand variability is presented. It allocates resources to all the departments such that it synchronizes departments through TTM and balance the patient flow in the OPC. The third section includes materials and methods that include data collection and analysis, model development, experimental design and simulation study. In the fourth section, the results from different scenarios created on real case studies are reported. Discussion and conclusions are presented in the fifth and sixth sections, respectively.

Predictive Resource Planning Model

The proposed RPM aims to match resources with demand in short term that is hour by hour so that this planning helps OPC to be maximally prepared to handle patient flow. That is, the RPM gives a margin to keep the OPC working, even when the reality is different from the plan. The predictive resource planning uses three levels of control (Figure 3).

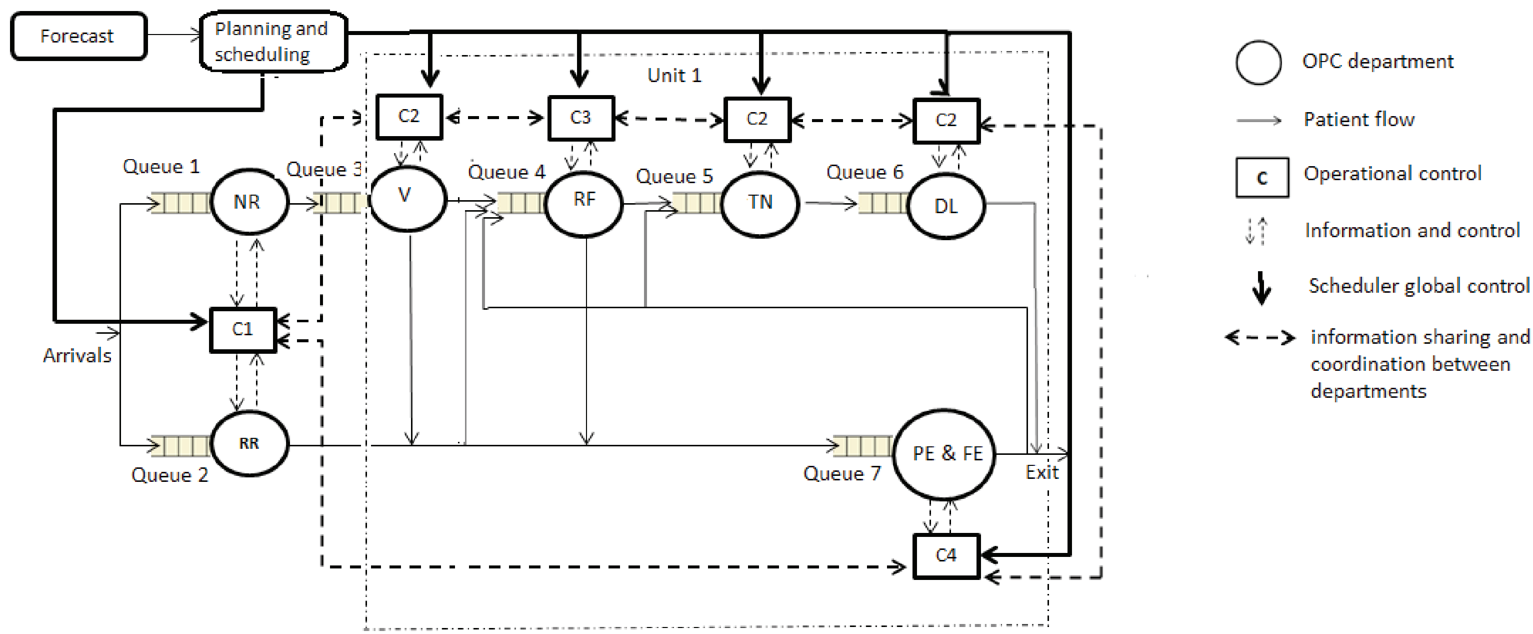

Figure 3. Functional Structure and Operational Control of Predictive Resource Planning and Scheduling in OPC System

Source: Derived from proposed resource planning model for OPC in AEH.

Box 2. Rules Proposed by Predictive RPM to OPC

Rule 3: Resource planning and scheduling time window is everyday/one day.
Rule 4: Resources are scheduled every hour.
Rule 5: Resources are scheduled by predictive planning model, and managers only follow the instructions.
Rule 9: Control of operations with respect to resources is organization-wide (OPC centric).

1. A forecast generator is used to forecast daily patient volume and patient arrival patterns. Patient volume is predicted based on the seasonal variations, such as vacations, short holidays and festivals, whereas patient arrival patterns are based on historical data.
2. A pace or Takt is set depending on the forecast data, in short term. A RPM identifies the resource requirements for all the departments throughout the day based on patient demand, considering constraints on resources and precedence in all the departments. This is organization-wide control and this resource plan is shared with all the departments.
3. Resources are planned to match the service rate with arrival rate, that is, resources are matched to patient demand.

To implement the predictive RPM, we proposed a few changes in the rules of the OPC (Box 2).

This model has proposed a modification of rules related to resources and their control but retains other rules 1, 2, 6 and 8 of the OPC system. The planning time window has been changed from a month to a day (Rule 3). Resources are scheduled every hour as compared to fixed over a day and month (Rule 4). The resource planning is performed by a planning and scheduling model, which incorporates variability in demand (Rule 5). This eliminates Rule 7 on corrective measures taken manually by managers. Control of operations remains local except for resource planning. Now managers in departments do not take planning and scheduling decisions related to resources but only follow the schedule given by the predictive planning model. And resources are scheduled in the OPC-wide view not department-centric.

In the proposed model, patient flow remains the same but departments have four states as compared to five states in Figure 2. The states are S1-initial state, S2-waiting state, S3-regular state and S4-finish state. The predictive planning model develops a resource plan based on demand forecast and schedule resources to all the departments accordingly. The model centrally determines organization-wide resource schedules, applied locally. There is no reactive state in the proposed model as in the current OPC (AEH) model: instead, the reaction time to follow the schedule becomes important. To make the planning better and schedule resources in organization-wide perspective, we use TTM in the predictive RPM.

Takt Time-based Predictive Resource Planning Model

Takt time in the OPC context can be translated as the average time at which a patient moves out of the OPC.

$$\text{Takt time} = \frac{\text{Effective available time in a day}}{\text{No.of patients serviced in a day}} \tag{1}$$

To understand TTM, we consider an example of a system with two departments A and B with mean service times of 10 ± 2 minutes and 15 ± 4 minutes, respectively. The demand is assumed to be 150 patients/day, and the system works for 12 hours. There are two breaks of 1 hour each. Therefore, the effective time will be 10 hours (600 minutes). The total cycle time will be 25 ± 6 minutes. (Here cycle time indicates time from start to end of service: some authors use it to indicate process time on one machine.) The Takt time will be (10 hours × 60 minutes)/150 = 4 minutes per patient. The Takt time of 4 minutes does not mean that the patients are treated in only 4 minutes (contradicting the service times in departments A and B), but it is every 4 minutes a patient should move out of the system. If the Takt time is less than 4 minutes then the service in the department is faster than the patient demand and the resources wait or are idle. If the Takt time exceeds 4 minutes, then the patient waits. To achieve a Takt of 4 minutes, the departments A and B need resources as calculated in Equation (2).

$$\text{Number of resources r} = \frac{\text{Service time}}{\text{Takt time}} \tag{2}$$

The stochasticity associated with service times is discussed in detail later in this Section. In this example, we use only mean service time to calculate a number of resources. Therefore, department A will need $10/4 = 2.5 \approx 3$ resources, and department B will need $15/4 = 3.75 \approx 4$ resources to keep the pace of the system. A system would work smoothly with this number of resources if the patient demand were stable or uniform. However, patient arrivals in OPCs are random and uncertain. During peak hours, patient demand shoots up. In the rest of the day, it is less but still, varies. Therefore, setting a single Takt time for a day does not capture the problems associated with variability and uncertainty in shorter time horizon like hours in a day. So, working hours in a day are divided into '*m*' time slots. Patient arrivals are forecast using historical data. The patient arrival times are generated and the number of patients per time slot is identified. Further, the Takt time per time slot is determined using Equation (1). Takt time remains fixed for a time slot but varies between time slots depending on demand, whereas cycle time remains same. By setting pace (Takt) in each time slot, we match service rate (with respect to departments) or throughput rate (with respect to the OPC system) with patient demand throughout the day. In this way, Takt time gives a real target for improvement.

The Takt time design considers demand variability as well as process variability. The effective time taken to do the task is service time $s(t)$. Occasionally the performance of the task is interrupted by a problem, and this occurs during the performance of the task with a probability p. Some amount of time (referred as 'surplus time' $\mathrm{sp}(t)$) is required to solve this problem (Hopp & Spearman, 1996; Smith & Tan, 2013). Therefore, the overall processing time until the task is completed is $p(t) = s(t) + p\ \mathrm{sp}(t)$. For example, when a certain patient at a department requires more than average service time, the excess time will not be won back by using less time with another patient. So the surplus time above the average service time should be included in the Takt time design. The squared coefficient of variation of service times (Equation (3)) includes the effects of surplus time, set-up times, irregularities, etc. The mean processing time is $E[p(t)] = E[s(t)] + pE[\mathrm{sp}(t)]$. Variance of $p(t)$ is var $s(t) + p\mathrm{varsp}(t) + p(1-p)(E[\mathrm{sp}(t)])^2$. The squared coefficient of variation of processing time $p(t)$,

$$C^2_{p(t)} = \frac{\mathrm{var\ s(t)} + p\mathrm{var\ sp(t)} + p(1-p)(\mathrm{E[sp(t)]})^2}{(\mathrm{E[s(t)]} + p\mathrm{E[sp(t)]})^2} \tag{3}$$

Mean service time and coefficient of variation are two fundamental process parameters with respect to cycle time performance (Jacobs et al., 2003). The waiting time is directly proportional to the coefficient of variation. Hence, to improve cycle time, it is important to minimize the coefficient of variation of service times, that is to minimize surplus time. Reducing process time variability is equivalent to increasing system capacity when measured by cycle time response (Curry & Feldman, 2010, pp. 109–123). The demand variability and process variability are buffered by first in, first out queue and buffer size (the number of patients that can wait in a queue) is estimated using Equation (4).

$$\text{Buffer size} = \frac{\mathrm{s(t)}_{\max} - \mathrm{s(t)}}{\mathrm{Takt(t)}} * \text{Patient Demand} \tag{4}$$

where $s(t)$ is the mean service time at preceding department and $s(t)_{max}$ is the maximum service time at preceding department. Each department could have patients being serviced and waiting (buffers) that is WIP. When WIP is maximum, demand in manufacturing plants could be blocked, but in AEH when demand exceeds buffer capacity or departmental capacity, then neither can patient demand be blocked (due to maximum WIP) nor patients be made to wait. Therefore, the total amount of work (patient demand) needs to be spread and synchronized along the departments in the OPC. To balance the patient flow (line balancing), resource planning should also consider the relation or ratio of service times between the departments.

As discussed there is the difference between forecasting and reality, the OPC system when working according to a particular Takt should offer a margin when operational problems occur due to variability (Rother, 2009). Operator load (total work an employee can perform) varies due to working speed of employees, and thus, depends on their skill. A realistic load should be given to the operator for smooth and efficient flow without jeopardising quality. This we take into account by multiplying Takt time by a factor called fudge factor or efficiency factor. Generally, an efficiency factor of 85 per cent to 95 per cent is used in designing industrial processes and allows operators to work at a productive rate. This results in properly built, quality products (Duggan, 2002; Fekete & Hulvej, 2013; Ortiz, 2006). In this case, we estimate it by coefficient of variation of the cycle time, $c(t)$.

$$f = CV_{c(t)} = \frac{\textit{Standard deviation of } c(t)}{\textit{Mean of } c(t)} \tag{5}$$

If OPC runs f% less than Takt time then it requires more resources, but it captures the operator load. If OPC runs f% more than Takt time, it captures constraints and allows buffers. Determining the Takt for a system is a design parameter that has to be assessed based on demand and service variability (buffer), resource constraints and space (layout) constraints. What is important is that the OPC systems understand the difference from Takt to target (throughput) as waste and should seek to improve it, not just accept it. Hence, the use of Takt time goes beyond the numerical calculations. The demand variability, process time variability like surplus time, set-up times, irregularities and fudge factor help in setting the margin to keep the OPC working when reality differs from planned.

We derive a new operations management where resources' planning is demand-driven and organization-wide, to keep all departments in pace, neither slow nor fast. On contrary to fast food chains, here resources are scheduled based on predicted demand by taking short-term variability (nearly real time) into account with an organization-wide perspective (not department-level but OPC-wide). We develop a mathematical model to identify a number of resources required, and the parameters used are listed in Table 1. The margin is a variable that is not directly used in the optimization problem. Reaction time is an exogenous variable that gets affected by the internal and external environment of the OPC. So its effect on dependent variables is analysed in the simulation model.

Table 1. List of Notations used in the Mathematical Model

Notation	Description
n	Number of departments in the OPC where n = 1, …, N
m	Number of time slots in a day where m = 1, …, M
$r_{m,n}$	Number of resources in mth time slot in nth department
R_T	Total number of available resources in the OPC
f	Coefficient of variation of cycle time $c(t)$ to set margin
Takt (t)	Takt time
z_m	Reciprocal of Takt time during mth time slot, $z = 1/\text{Takt}(t)_m(1+f)$
g_n	Ratio of Average Service Time of nth Department to Average Cycle Time
h_n	Ratio $s(t)_n/s(t)_{n+1}$
$s(t)_n$	Mean service time of the nth department
$p(t)_n$	Processing time that includes service time and surplus time of nth department
$C^2_{p(t),n}$	Squared coefficient of variation of processing time $p(t)$ of nth department

Source: Authors' own.

The optimization problem is to find the minimum number of resources to achieve the required Takt time. That is we have to find

$$\text{Min}\sum_{m=1}^{M}\sum_{n=1}^{N} z_m r_{m,n} \tag{6}$$

subject to the condition that

$$\sum_{m=1}^{M}\sum_{n=1}^{N} g_n r_{m,n} \le R_T \tag{7}$$

$$\sum_{n=1}^{N} r_n \le h_n r_{n+1} \tag{8}$$

$$\sum_{m=1}^{M} \frac{r_{m,n}}{s(t)_n(1+C^2_{p(t),n})} \le z_m \tag{9}$$

$$\sum_{n=1}^{N} r_n \le R_T \tag{10}$$

$$r_{m,n} \ge 1 \tag{11}$$

$$r_{m,n}\,\text{integer} \tag{12}$$

The objective (6) minimizes the number of required resources in each department in each time slot, to achieve the required Takt time. Resources are the decision variables in this optimization model.

Constraints (7) and (8) deal with the same or identical service rate in all the departments, neither too fast nor too slow, but in pace and within the total available resources (organization-wide optimization). Constraint (9) relates the service variability, cycle time and Takt time to the resources. The solution is not just increasing resources when departmental demand is high, or conversely, but doing so without disturbing the flow of successive upstream or downstream departments. This synchronizes spread of work between departments in the OPC. Constraint (10) shows that the number of resources in all the departments in the *m*th time slot should be within an available total. This constraint takes care of the criterion of line balancing and the buffers. Constraint (11) assures that at least one resource is always allocated to every department in all the time slots. Finally, constraint (12) requires the integer assignment of resources.

The mathematical model was solved using ILP. The ILP output was the resource plan that indicated the number of resources in each department in each of the time slots. We explain the planning model with an example. We selected 11-time slots of an hour each (based on demand analysis). Patient demand was identified and Takt time was determined for each time slot. The resources were scheduled based on this predictive plan for both the units as shown in Table 2. These resource plans were used in the OPC simulation model to see its effect on performance measures. The time slots where patient demand is high have smaller Takt time. It indicates the system needs to work faster and for this, the required number of resources must be scheduled to control waiting times. In time slots where patient demand is less, the number of resources scheduled can be reduced and under-utilization can be avoided.

Table 2. Output of the Mathematical Model Compared to the Existing Resource Schedule

Time Slots	Takt Time in Mins	New Registration	Review Registration	Vision 1	Preliminary and Final Exam 1	Refraction 1	Tension 1	Dilatation 1	No. of Resources in Unit 1/ Both Units
1	0.8	2	2	1	4	4	2	1	16/28
2	1.2	2	2	1	4	4	2	1	16/28
3	0.6	3	3	2	4	4	2	1	20/32
4	0.7	2	3	1	4	4	2	1	17/29
5	0.7	2	3	1	4	4	2	1	17/29
6	1.2	1	2	1	3	4	1	1	13/23
7	1.4	1	2	1	3	3	1	1	12/21
8	2.4	1	1	1	2	2	1	1	9/16
9	5.4	1	1	1	1	1	1	1	7/12
10	12	1	1	1	1	1	1	1	7/12
11	20	1	1	1	1	1	1	1	7/12
Currently used schedule for whole day	—	2	2	1	3	5	2	1	16/28

Source: Obtained from resource plan of the OPC in AEH (optimization model-Matlab).
Note: The OPC currently follows the resource schedule as shown in the last row of Table 2. As per current scheduling, the total number of resources scheduled in both units is 28, only unit 1 and registration is 16, and this remains the same for a month.

Materials and Methods

Data Collection and Analysis

The initial data were collected by observations and interviews of patients and hospital staff (ophthalmologists, paramedical, managers and administrative) of AEH. Patient and process data of 6 months were obtained from the in-house software: Integrated Hospital Management System (IHMS) and Clinical Management System (CMS). The data collected included the patient volume, arrival times and exit time, service times, resource schedule, waiting times and cycle times (which include registration time, service time and waiting time in all departments), and the reaction time, obtained through interviews of staffs and managers. Data of 53,802 patients were analysed, and the data-fitting tool Easy Fit was used to determine the probability distribution of service time and patient arrival time. The data analysis showed that the patient arrival pattern had two peaks, at around 8:00 AM and 10:00 AM. Therefore, a bimodal Poisson distribution (Karlis & Xekalaki, 2005; Li & Zha, 2006) was selected to generate the arrival times (Equation (13)).

$$P = \{v_1, v_2\} \text{ and } \lambda = \{\lambda_1, \lambda_2, P\} \tag{13}$$

where P is the sum of two Poisson distributions with mean arrivals λ_1 and λ_2 mixed with proportions $v_1 = 0.35$ and $v_2 = 0.65$. The goodness of fit test for input and output distribution was conducted using the Kolmogorov–Smirnov test. The workflow of the model was verified using flowcharts and a structured walk-through by the managers of the OPC.

Model Development

A discrete event simulation model of the OPC was developed using Java. A patient was an entity whose progress was tracked. Service times were uniformly distributed between the minimum and maximum service times from empirical data for each department and were randomly generated. The patient arrival time and the number of resources in each department were used from the empirical data. The managers of the OPC verified the program. Further, to improve the accuracy of the simulation model, it was calibrated by assigning the reaction time randomly between 20–30 minutes. The simulation model was run with the empirical data and performance parameters were collected. The results from the simulation model were compared with the empirical data of the OPC for validation as shown in Table 3, and there was no statistical difference between the two.

Table 3. Validation of the Simulation Model with the Existing OPC in AEH

Patient Demand	Waiting Time in Minutes			Cycle Time in Minutes		
	Mean ± SD			Mean ± SD		
	Existing AEH	Simulation Model	*p*-value	Existing AEH	Simulation Model	*p*-value
Low	48.6 ± 12.45	45.7 ± 10.57	0.5	98.9 ± 14.25	96.2 ± 12.54	0.4
Medium	68.2 ± 18.56	66.1 ± 19.11	0.4	122.3 ± 17.83	119.9 ± 20.43	0.3
High	82.1 ± 25.02	79 ± 23.54	0.6	138.9 ± 27.12	137.9 ± 25.32	0.5

Source: AEH data from IHMS and CMS.
Note: SD—Standard deviation.

The difference in waiting times and cycle times of the simulation model and existing AEH is due to local optimization performed by paramedical staffs in the units. In the existing AEH, whenever congestion is observed paramedical staffs manually change the sequence for patients in RF and PE. But in the simulation model, sequencing is performed for each patient.

Experimental Design

It is observed from literature and case study in AEH that the difference between planning and reality results in operational problems. In this study, we measured waiting times and cycle times. This experimental design had three factors: patient demand, scheduling rules (control of operations) and reaction time. Patient demand in AEH is huge and variable and affects the waiting times. In AEH, the average demand is high around 1800 patients/day with 30.8 per cent of the monthly patient demand being 1000–1600 patients/day, 49.9 per cent 1600–2000 patients/day and 19.3 per cent being greater than 2000 patients/day. Therefore, we classified patient demand into low, medium and high. The scheduling rules varied in two levels: existing with fixed number of resources (local) and with predictive resource plan based scheduling with varying number of resources within a day (organization-wide). Since the reaction time affects the waiting times, different reaction times in minutes: $r(t1)$ that is ≤ 10, $11 \leq r(t2) \leq 20$ and $21 \leq r(t3) \leq 30$ were used for the experiment. Reaction times are randomly assigned to the departments in the selected range. Additionally, to analyse the effect of reaction times on departmental performance, we selected six combinations of reaction times based on service times (high and low). There were in total $2^1 \times 3^1 \times 9^1 = 54$ experiments and the performance measures were recorded for all the experiments in the design. A full factorial experiment was carried out to estimate the effect of selected factors on performance parameters.

Simulation Runs

The experimental design has been replicated 10 times with 540 runs to estimate the variability associated with the phenomenon. The simulation of a day took around 3–4 minutes per day. The seed in random variate was varied to generate different arrival times for the same mean patient arrivals. The same randomizer input was used for simulation, with the two different scheduling scenarios: existing (fixed) and predictive plan-based (proposed) resource scheduling. This assured the results obtained were not due to randomness. The mean and standard deviation of the waiting time and cycle time were collected. These results were compared with the existing AEH. ANOVA-tests were conducted for statistical comparisons at a significance level of 0.05. Additionally, ANOVA-tests were performed using IBM SPSS to determine the significance of main effects and interaction effects of predictive planning and reaction time on waiting times and cycle times. Further, the proposed model was implemented in units 1 and 2 of the OPC. The daily resource schedule based on predictive planning was implemented for a month, the performance measures were collected from the IHMS, and CMS are presented in the results section.

Results

The mean cycle time for existing scheduling scheme and predictive resource planning (simulation) was 120.1 ± 19.7 minutes and 89.3 ± 9.3 minutes, respectively. The mean waiting time was 66.3 ± 18.7 minutes and 37.5 ± 8.9 minutes, respectively. The mean and standard deviation of the waiting time and cycle time for existing and predictive planning are compared in Tables 4 and 5, respectively.

Table 4. Comparison of Average Waiting Time in Minutes of Existing and Predictive Plan-based Scheduling

Patient Demand/ Reaction Time $r(t)$	Existing Scheduling with Local Control of Operations			Predictive Plan-based Scheduling with Organization-wide Control of Operations		
	Mean ± SD			Mean ± SD		
	$r(t1)$ min	$r(t2)$ min	$r(t3)$ min	$r(t1)$ min	$r(t2)$ min	$r(t3)$ min
Low	44.3 ± 8.3	49.3 ± 8.6	52.1 ± 10.9	15.1 ± 6.7	32.6 ± 8.1	44.7 ± 9.2
Medium	56.6 ± 15.8	62.9 ± 15.7	71.8 ± 18.2	24.1 ± 7.3	36.2 ± 9.2	52.9 ± 10.9
High	69.5 ± 20.1	75.3 ± 21.9	83.1 ± 23.9	30.1 ± 7.8	44.6 ± 9.9	54.2 ± 11.1

Source: Derived from AEH data and simulation model.
Note: SD—Standard deviation.

Table 5. Comparison of Average Cycle Time in Minutes of Existing and Predictive Plan-based Scheduling

Patient Demand/ Reaction Time $r(t)$	Existing Scheduling with Local Control of Operations			Predictive Plan-based Scheduling with Organization-wide Control of Operations		
	Mean ± SD			Mean ± SD		
	$r(t1)$ min	$r(t2)$ min	$r(t3)$ min	$r(t1)$ min	$r(t2)$ min	$r(t3)$ min
Low	93.5 ± 8.4	99.6 ± 8.8	101.1 ± 10.8	70.3 ± 6.9	85.8 ± 8	101.6 ± 9.3
Medium	107.5 ± 16.1	112.3 ± 16.9	119.9 ± 19.9	74.8 ± 7.7	86.4 ± 9.7	104.8 ± 12.3
High	119.7 ± 20.9	121.1 ± 22.1	132.7 ± 24.5	78.2 ± 7.5	94.5 ± 10.2	107.1 ± 12.4

Source: Derived from AEH data and simulation model.
Note: SD—Standard deviation.

The average waiting time was found to be reduced and regulated in all departments. The waiting times in both registrations, preliminary and final examinations have been reduced significantly. At the same time, we notice that there is an increase in average waiting time in the vision department by a few minutes as shown in Figure 4. The figure also compares the waiting times in departments for various combinations of different reaction times and their respective cycle times. Reaction times are selected based on low and high service times. Example for mix1: departments with low (L) service times (<5 minutes) is $r(t1)$ and high (H) service times is $r(t2)$.

Reaction time is an exogenous variable that influences performance measures. The two-way ANOVA-tests ($p = 0.05$) showed the significance of main effects and interaction effects of predictive planning and reaction time are presented in Table 6.

Planning has a significant main effect on waiting times, whereas the main effect of reaction time is not significant. The interaction effects of planning and reaction time are significant. Further, pairwise comparisons were performed for the three levels of reaction times: $r(t1) - r(t2)$, $r(t2) - r(t3)$ and $r(t1) - r(t3)$ and their significance were 0.205, 0.226 and 0.023, respectively. The interaction effect of $r(t1)$ and $r(t2)$ was comparatively more than $r(t3)$ (Figure 5). The results show that reaction time contributes around 5 per cent (4.89 per cent) in reducing waiting times. As reaction times between departments varied, tests of between-subjects effects were conducted. Main and interaction effects of departments, planning and reaction time on performance measures were analysed (with p-values for department × planning (0.001), department × reaction time (0.50) and planning × reaction time (0.003)). We observe that main and interaction effects of planning are significant in all the departments, whereas reaction time is not significant in departments with shorter service times.

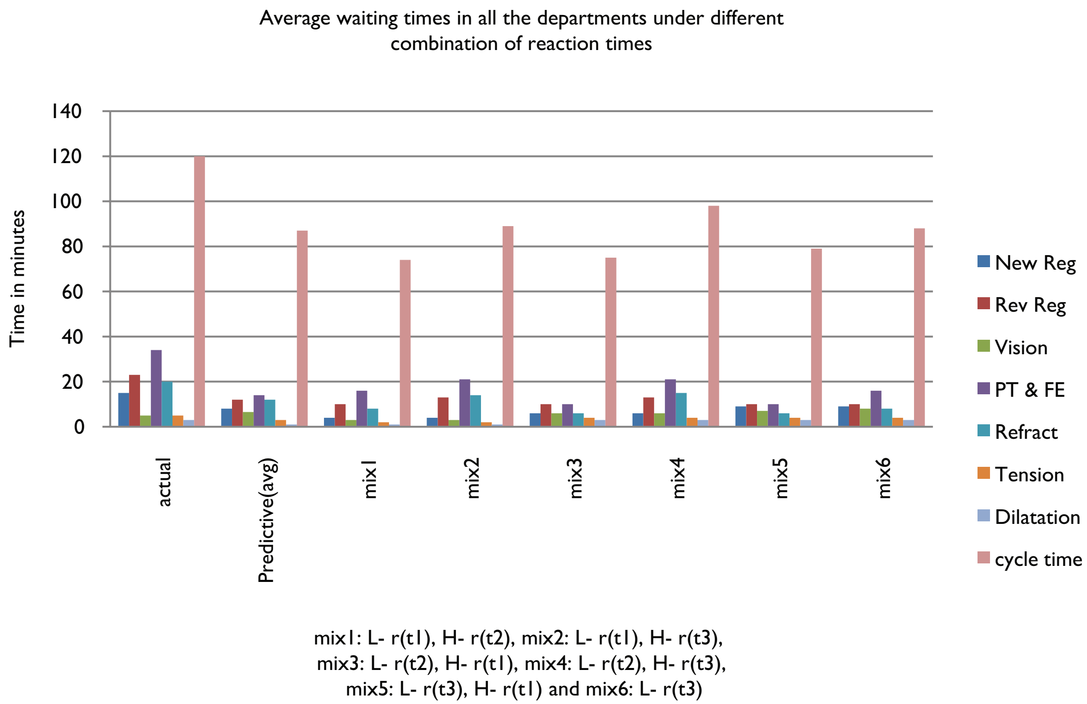

Figure 4. Effect of Reaction Times on Average Waiting Times in All Departments and Average Cycle Time

Source: Derived from AEH data and resource planning model.

Table 6. Tests of Between-subjects Effects

Source	Type III Sum of Squares	df	Mean Square	*F*	Significance
Intercept	829,177.965	1	82,9177.965	46.410	0.021
Planning	52,486.767	1	52,486.767	216.089	0.005
Reaction time	35,732.657	2	17,866.328	73.556	0.068
Planning*Reaction time	485.788	2	242.894	3.274	0.049

Source: Derived from minitab data.

The predictive resource planning was implemented for a month in OPC of AEH. The performance measures were collected from IHMS and CMS. The mean cycle time and waiting time after implementation were 92.4 ± 9.1 minutes and 39.0 ± 8.2 minutes, respectively (Table 7). The existing resource plan had a fixed number (28) of resources in both units in a day whereas the predictive resource plan has a minimum of 12 to a maximum around 30–34 resources. The additional resources were scheduled from different clinics of AEH and the resources that were reduced during some time slots were utilized for maintaining patient records (back-end work). As seen in Table 2, when compared to the existing schedule, the number of resources required during time slots 3 and 4 is higher and in few time slots, it is lower.

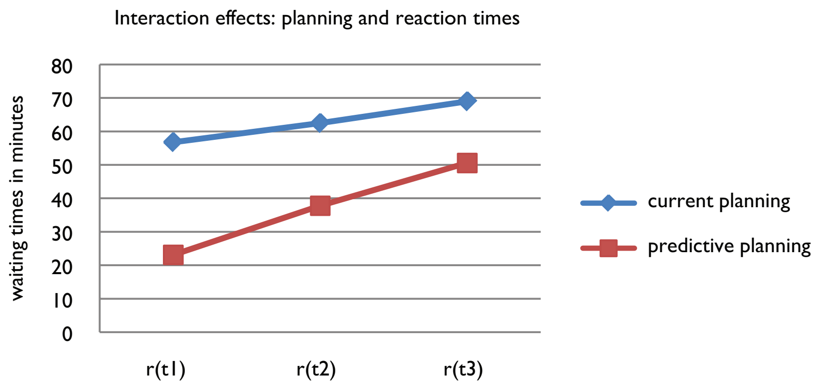

Figure 5. Result of Two-way ANOVA Test

Source: Derived from minitab data.

Table 7. Average Waiting Times and Cycle Times Before and After Implementation of the Predictive Resource Planning in the OPC

Patient Demand	Number of Resources			Waiting Time in Minutes			Cycle Time in Minutes		
	Before	After		Before	After		Before	After	
		Min	Max	Mean ± SD	Mean ± SD	*p*-value	Mean ± SD	Mean ± SD	p-value
Low	28	12	30	48.6 ± 12.5	34.3 ± 7.1	0.03	96.0 ± 14.3	85.8 ± 7.7	0.04
Medium	28	12	32	68.2 ± 18.6	39.3 ± 8.4	0.01	118.3 ± 17.8	92.4 ± 8.9	0.02
High	28	12	34	82.1 ± 25.0	43.4 ± 9.1	0.01	136.0 ± 27.1	99.1 ± 10.6	0.02

Source: Derived from AEH data before and after implementation from IHMS and CMS.
Note: SD—Standard deviation.

The performance measures of the proposed model were significantly different from those of the existing situation. A *p*-value of 0.05 was selected, meaning that for a *p*-value less than 0.05, the null hypothesis is rejected, and the difference is statistically significant.

Discussion

The OPC system in AEH consists of disparate departments that schedule their resources departmentally and ahead of time (once a month). The OPC being an open loop system is prone to variability and the current method of planning caused the formation of bottlenecks in a few departments and resulted in prolonged waiting times and cycle times. The OPC in AEH like other hospitals (as seen in literature) managed its services (by planning resources locally) based on average demand (day-wise) not the actual demand and its variability.

In this study, resources were planned and scheduled with an organization-wide perspective based on the actual demand (near to real time) by incorporating short-term variability. Our findings confirm that the predictive planning model reduces the average waiting time by 43.4 per cent during simulation (from 66.3 minutes to 37.5 minutes) and by 41.1 per cent during its implementation (from 66.3 minutes to 39.0 minutes) in the OPC. Also, standard deviations of waiting times were reduced significantly. The study demonstrates the effect of patient demand, planning and scheduling rules, control of operations and reaction times on waiting times and cycle times (Table 5). It is found that organization-wide control of operations in departments for planning and scheduling resources improves waiting times compared to the local control of operations.

Besides scheduling rules and control of operations, the reaction time, an exogenous variable also influences waiting time to some extent. ANOVA was used to determine the main and interaction effects of predictive planning and reaction time. Effects of predictive planning are significant in all departments on performance measures. Improvement in reaction times alone does not influence waiting times greatly, but along with predictive planning, has a significant effect on waiting times as well as cycle times (Table 6). The reaction time varies between departments and the reaction times of the departments that have longer service times have a significant influence on waiting times and cycle times (Figure 4). During high patient demand, the influence of reaction time on performance measures is of greater importance and being prepared for variability improves waiting times.

Waiting times have increased in some departments like vision by a few minutes. This is because of constraint (7) that balances the patient flow within the available resources. The number of resources required to achieve a particular Takt time may be high, but due to resource and flow balance constraints (a spread of work equally), the scheduling model reduces the number of resources in few departments in the patient flow. Because of this, waiting times in few of these departments are increased (see Figure 4).

The resource schedule obtained by the proposed model shows that a different number of resources are scheduled during different times of the day (Table 2). When patient demand is varying, why should the service rate be constant? So OPCs need to plan their resources according to demand and for that OPC should have resource flexibility. This predictive planning model takes into account the variability in demand in short term (hour by hour) and makes planning better, near to real time. The current planning model estimates the total amount of work within a day and adapts resources accordingly. Planning is a decision taken much ahead of time and when the reality is different from planning, departments optimize locally leading to sub-optimization. The predictive resource planning considers real-time operational problems that occur due to variability and uncertainty, during planning itself. Additionally, this planning model helps the management to identify resource requirements: types of resources and number of resources, for the future demand growth and expansion.

Takt time management synchronizes total amount of work and spread of work. It has been observed that it is applied to systems with even and stable demand, for longer time horizon and with an organization-wide perspective. But in this study, TTM has been applied to uneven and unstable demand (all walk-in patients), in a shorter time horizon. There is no control on patient demand that is much higher than the maximum WIP the OPC can manage and patient demand cannot be blocked as done in a manufacturing system. Therefore, setting the Takt over a shorter period captured variability in short-term (one hour) and resources were scheduled every hour. The variability and uncertainty are inherent to OPC systems and cannot be eliminated. But they can be better handled than by attempts to eliminate them. Using Takt time in resource planning brings predictability into the design effort and eliminates the unplanned overtimes of the resources. This implies that variability needs to be accounted for during planning for real-time workflow optimization. It balances patient flow by matching the supply in accordance with the patient demand. The study demonstrates that TTM can also be applied to open loop systems where variability is high.

The literature shows that patient demand has been matched with resources, but planning often considered forecasting only over a long period. Planning often failed to incorporate variability in the short term and this increased the gap between planning and reality. In this study, the predictive RPM integrates stochasticity in patient demand throughout the day and matches the resources accordingly in short term. It also incorporates the interdependencies between departments and generates a plan that optimizes resources in OPC perspective (organization-wide).

There were some factors that caused the implementation results to be different from what could be expected on the basis of the simulation study. Though the resources were scheduled based on predictive planning model, sometimes they arrived late to the departments as they were shared between other clinics in AEH. The predictive planning was implemented in only two units of the OPC but other clinics followed their regular scheduling rules. The planning model considered the constraints on resources and space/layout but did not consider resource sharing. The late arrival of resources affects the waiting times. The non-uniform patient arrivals or patient distribution within the time slots affects the waiting times. Sometimes, more patients arrive at the end of a time slot and this increase the waiting time.

Healthcare settings like OPCs or hospitals vary as they differ in their complexity, patient groups and processes. Therefore, requirements for planning might vary. Many processes and arrivals are not deterministic and need optimization in the short term through advanced planning systems. But, how we use this planning system might depend not only on the algorithms but also on factors such as the organization of the hospital/OPC, cross-skills of the staff so that they can be transferred between departments and the layout of the building. As future work, we intend to analyse how TTM could influence resource utilization and how to extend this planning model incorporating resource sharing.

Conclusion

This study shows how predictive resource planning (near to real time) improves waiting times and cycle times in OPCs. This study integrates two facets of planning: demand-driven and an organization-wide perspective. The resource planning is based on actual (near to real time) demand. Short-term variability in demand should be incorporated to make planning better and near to reality. The planning with organization-wide perspective overcomes the problem of unregulated waiting times, that is departments work neither too fast not too slow. Predictive RPM utilized TTM to set pace between demand and service, and balance patient flow by organization-wide optimization.

Resource planning in open loop systems like OPC systems is especially important in operations management as waiting times are a major concern of quality care. The variability and uncertainty in OPC systems can be minimized (nearly closed loop) by incorporating patient volume, patient types, resources and reaction times in real time optimization, thus improving the waiting times and cycle times. Although, this reaction time is an exogenous variable the predictive resource planning helps OPCs to avoid longer reaction times as the demand-driven resource schedule is known in advance. It was seen from this study that the model made a positive impact with some of the drawbacks mentioned above and even with reaction times as long as 30 minutes. With shorter reaction times, the model performs better. As the model was robust, small drawbacks in the implementation did not alter outcomes. This planning model has been implemented to eye care OPC and can probably be extended to general hospitals by considering their demand forecast, precedence constraints and workflow complexities.

References

Activity-Report. (2014–2015). *Aravind eye care system* (p. 76). Madurai. Retrieved 8 April 2017, from http://www.aravind.org/default/researchnewcontent/annualreports

Alvarez, Roberto dos Reis, & Antunes, José Antonio Valle, Jr. (2001). Takt-time: Concepts and context in Toyota Production System. *Gestão & Produção*, *8*(1), 1–18.

Andersen, Michael Moesgaard, & Poulfelt, Flemming. (2014). *Beyond strategy: The impact of next generation companies*. New York, NY: Routledge.

Brilliant, Larry, & Brilliant, Girija. (2007). Aravind: Partner and social science innovator (Innovations case discussion: Aravind eye care system). *Innovations: Technology*, *Governance*, *Globalization*, *2*(4), 50–52.

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

Chan, H. Y., Lo, S. M., Lee, L. L. Y., Lo, W. Y. L., Yu, W. C., Wu, Y. F., … & Chan, J. T. S. (2014). Lean techniques for the improvement of patients' flow in emergency department. *World Journal of Emergency Medicine*, *5*(1), 24–28. DOI: 10.5847/wjem.j.issn.1920-8642.2014.01.004

Chaudhary, Bhupinder, Modi, Ashwin G., & Reddy, Kalyan. (2012). Right to sight: A management case study on Aravind Eye Hospitals. *ZENITH International Journal of Multidisciplinary Research*, *2*(1), 447–457.

Curry, Guy L., & Feldman, Richard M. (2010). *Manufacturing systems modeling and analysis* (2nd ed.). Berlin; Heidelberg: Springer-Verlag.

Day, Robert W., Dean, Matthew D., Garfinkel, Robert, & Thompson, Steven. (2010). Improving patient flow in a hospital through dynamic allocation of cardiac diagnostic testing time slots. *Decision Support Systems*, *49*, 463–473. DOI: 10.1016/j.dss.2010.05.007

Duggan, Kevin J. (2002). *Creating mixed model value steams—Practical lean techniques for building to demand*. Boca Raton, FL: Productivity Press.

Edward, G. M., Das, S. F., Elkhuizen, S. G., Bakker, P. J. M., Hontelez, J. A. M., Hollmann, M. W., … & Lemaire, L. C. (2008). Simulation to analyse planning difficulties at the preoperative assessment clinic. *British Journal of Anaesthesia*, *100*(2), 195–202.

Eswaramoorthi, M., Kathiresan, G. R., Jayasudhan, T. J., Prasad, P. S. S., & Mohanram, P. V. (2012). Flow index based line balancing: A tool to improve the leanness of assembly line design. *International Journal of Production Research*, *50*(12), 3345–3358. DOI: 10.1080/00207543.2011.575895

Fekete, Milan, & Hulvej, Jaroslav. (2013). *"Humanizing" Takt time and productivity in the labor-intensive manufacturing systems*. Paper presented at the Knowledge Management and Innovation Management, Knowledge and Learning International Conference, Croatia.

Gupta, Diwakar, & Denton, Brian. (2008). Appointment scheduling in health care: Challenges and opportunities. *IIE Transactions*, *40*, 800–819.

Harper, Paul R. (2002). A framework for operational modelling of hospital resources. *Health Care Management Science*, *5*(3), 165–173.

Hassan, M. H., Zaqhloul, A. A., & Mokhtar, S. A. (2005). The probability distribution of attendance to hospital emergency units for school students in Alexandria. *The Journal of the Egyptian Public Health Association*, *80*(1–2), 127–151.

Hopp, Wallace J., & Lovejoy, William S. (2012). *Hospital operations: Principles of high efficiency health care* (*FT Press Operations Management*). New Jersey: FT Press.

Hopp, Wallace J., & Spearman, Mark L. (1996). *Factory physics: Foundations of manufacturing management*. Irwin: The McGraw-Hill.

Huang, X. M. (1994). Patient attitude towards waiting in an outpatient clinic and its applications. *Health Services Management Research: An official Journal of the Association of University Programs in Health Administration* / HSMC, AUPHA, *7*(1), 2–8.

Jacobs, J. H., Etman, L. F. P., van Campen, E. J. J., & Rooda, J. E. (2003). Characterization of operational time variability using effective process times. *IEEE Transactions on Semiconductor Manufacturing*, *16*(3), 511–520.

Karlis, Dimitris, & Xekalaki, Evdokia. (2005). Mixed Poisson distributions. *International Statistical Review, 73*, 35–58. Retrieved from http://www.jstor.org/stable/25472639

Li, Jia, & Zha, Hongyuan. (2006). Two-way Poisson mixture models for simultaneous document classification and word clustering. *Computational Statistics and Data Analysis*, *50*(1), 163–180. DOI: 10.1016/j.csda.2004.07.013

Liker, J. K. (2004). *The Toyota way: 14 management principles from the world's greatest manufacturer*. New York, NY: McGraw-Hill.

Ludwig, Martijn, Merode, Frits Van, & Groot, Wim. (2010). Principal agent relationships and the efficiency of hospitals. *European Journal of Health Economics, 11*(3), 291–304. DOI: 10.1007/s10198-009-0176-z

Mansdorf, Bruce D. (1975). Allocation of resources for ambulatory care—A staffing model for outpatient clinics. *Public Health Reports*, *90*(5), 393–401.

Mehta, Pavithra K., & Shenoy, Suchitra. (2011). *Infinite vision—How Aravind became the greatest business case for compassion*. San Franciso: Berrett-Koehler.

Molema, J. J. W., Groothuis, S., Baars, I. J., Kleinschiphorst, M., Leers, E. G. E., Hasman, A., & van Merode, G. G. (2007). Healthcare system design and part time working doctors. *Health Care Management Science*, *10*, 365–371. DOI: 10.1007/s10729-007-9032-9

Natchiar, G., Thulasiraj, R. D., & Sundaram, R. Meenakshi. (2008). Cataract surgery at Aravind Eye Hospitals: 1988–2008. *Community Eye Health*, *21*(67), 40–42.

Nguyen, T. B. T., Sivakumar, A. I., & Graves, S. C. (2015). A network flow approach for tactical resource planning in outpatient clinics. *Health Care Management Science*, *18*(2), 124–136. DOI: 10.1007/s10729-014-9284-0

Ortiz, Chris A. (2006). *Kaizen assembly: Designing*, *constructing and managing a lean assembly line*. Boca Raton, FL: CRC Press.

Pillay, D. I., Ghazali, R. J., Manaf, N. H., Abdullah, A. H., Bakar, A. A., Salikin, F., ... & Ismail, W. I. (2011). Hospital waiting time: The forgotten premise of healthcare service delivery? *International Journal of Health Care Quality Assurance*, *24*(7), 506–522. DOI: 10.1108/09526861111160553

Rangan, V. K., & Thulasiraj, R. D. (2007). Making sight affordable (Innovations case narrative: Aravind eye care system). *Innovations: Technology*, *Governance*, *Globalization Fall*, *2*(4), 35–49.

Rother, Mike. (2009). *Toyota Kata: Managing people for improvement*, *adaptiveness and superior results*. New Delhi: McGraw-Hill Professional.

Sandanayake, Y. G., & Oduoza, C. F. (2009). Dynamic simulation for performance optimization in just-in-time-enabled manufacturing processes. *International Journal of Advanced Manufacturing Technology*, *42*(3–4), 372–380. DOI: 10.1007/s00170-008-1604-4

Smith, James MacGregor, & Tan, Baris. (2013). *Handbook of stochastic models and analysis of manufacturing system operations* (Vol. 192). New York, NY: Springer-Verlag.

van Merode, Godefridus G., Groothuis, Siebren, & Hasman, Arie. (2004). Enterprise resource planning for hospitals. *International Journal of Medical Informatics*, *73*(6), 493–501. DOI: 10.1016/j.ijmedinf.2004.02.007

Vermeulen, Ivan B., Bohte, Sander M., Elkhuizen, Sylvia G., Lameris, Han, Bakker, Piet J. M., & Poutre, Han La. (2009). Adaptive resource allocation for efficient patient scheduling. *Artificial Intelligence in Medicine*, *46*(1), 67–80.

Vissers, J. M. H., Bertrand, J. W. M., & de Vries, G. (2001). A framework for production control in health care organizations. *Production Planning & Control: The Management of Operations*, *12*(6), 591–604.

Voort, M. M. Rouppe van der, Merode, F. G. van, & Berden, B. H. (2010). Making sense of delays in outpatient specialty care: A system perspective. *Health Policy*, *97*(1), 44–52. DOI: 10.1016/j.healthpol.2010.02.013

Yurko, Lynne C., Coffee, Tammy L., Fusilero, Jane, Yowler, Charles J., Brandt, Christopher P., & Fratianne, Richard B. (2001). Management of an inpatient-outpatient clinic: An eight-year review. *Journal of Burn Care & Research*, *22*(3), 250–254.

Zhu, Zhecheng, Heng, BeeHoon, & Teow, KiokLiang. (2012). Analysis of factors causing long patient waiting time and clinic overtime in outpatient clinics. *Journal of Medical Systems*, *36*(2), 707–713. DOI: 10.1007/s10916-010-9538-4