Original Research

INQUIRY: The Journal of Health Care Organization, Provision, and Financing
Volume 56: 1–13

DOI: 10.1177/0046958019837430
journals.sagepub.com/home/inq
SAGE

Simulation of Quality Death Spirals based on Human Resources Dynamics

Ali Alizadeh-Zoeram[1], Alireza Pooya, PhD[1], Zahra Naji-Azimi, PhD[1], and Ali Vafaee-Najar, PhD[2]

Abstract

Although the hospital managers always try to improve the quality of the medical services, sometimes their efforts might affect reversely and push the system in what is so commonly called as "the death spirals of quality." The most important reason of falling into these spirals is the lack of a systemic thought that considers the feedback relationships between the numerous effective variables in the system performance, such as human resources service capacity. In this regard, the purpose of the present research is to design and simulate a dynamic human resources service capacity–based model to demonstrate the death spirals of quality phenomenon based on the service time per service and the possibility of error generation along with identifying the policies to cope with them. The system dynamics simulation approach is used to show the dynamics of the capacity of service from the standpoint of human resources. A model is simulated for the services of a hospital clinic as a case study. The simulation results of the designed dynamic model express that applying the desired policies for the case study can provide a good basis for fighting these spirals in a dynamic situation.

Keywords

health care, service quality, human resources, service capacity, system dynamics, simulation

Introduction

One of the main policies of service organizations such as medical service centers, for achieving competitive advantages and improving their performance, is concentrating on the "quality" subject. In this regard, those experts who used the product feature or technical quality approach to conceptualize the service quality focus on the standard of service performance.[1] On the contrary, quality for those experts with the operational or functional view is defined as a function of the operation or performance features such as waiting times, rate of errors in interactions, and processing times to control the process.[2-3] Lovelock and Wright[4] consider 7 quality gaps as the main determinants of the service quality including knowledge gap, standards gap, delivery gap, internal communication gap, perception gap, interpretation gap, and service gap.

Statistical reports show that although the quality of most of the manufactured products in recent decades have been improved, medical services quality index did not experience much of a change in 2015 compared with the reports of 1995 which was 74%. However, at some points of time, the value of the index has decreased. In other words, the quality erosion phenomenon has occurred in medical services (according to American Customer Satisfaction Index).[5] Despite the fact that hospital managers always try to improve the quality of medical services, based on the work by Senge and Oliva (1993), corrective actions in some systems may have reverse consequences.[6] Evaluating the quality improvement actions not only requires concentration on measuring the effectiveness but also needs the factors influencing their change and interactions to be identified.[7] A true example of such a claim in the field of service quality can be observed in the study by Oliva and Sterman. They, based on their studies in service organizations, came up with the death spirals phenomenon in which some of the actions of the organization aiming to improve its service quality eventually lead to forming some quality erosion loops or spirals and consequently its gradual death, which is the result of the interactions and feedbacks of different actions.[8] All of these evidence prove that in providing medical services, a complex and dynamic system exists, for which the lack of a systemic view to the designed

[1]Ferdowsi University of Mashhad, Mashhad, Iran
[2]Mashhad University of Medical Sciences, Mashhad, Iran

Received 29 March 2018; revised 30 December 2018; revised manuscript accepted 8 February 2019

Corresponding Author:
Alireza Pooya, Associate Professor of Management, Department of Management, Administrative and Economic Sciences Faculty, Ferdowsi University of Mashhad, Mashhad, Iran.
Email: alirezapooya@um.ac.ir

Table 1. Research works of Human Resources Service Capacity in Health Care.

Model	Model type	References
Linear programming model	Analytical	Willemain and Moore[18] and Smith et al[19]
Goal programming	Analytical	Kwak and Lee[20]
Integer programming	Analytical	Jaumard et al[21] and Mobasher[22]
Stochastic linear programming	Analytical	Lin et al[23]
Multiobjective integer programming	Analytical	Punnakitikashem et al[24]
Six Sigma	Analytical	Jayasinha[25]
Discrete event	Simulation	Rohleder et al,[26] Ghanes et al,[27] Gul et al,[28] Banks et al,[29] Borshchev and Filippov,[30] Acker et al,[31] Evans et al,[34] Duguay and Chetouane,[35] and Al-Araidah et al[36]

corrective actions for improving the service quality can lead to these death spirals of quality.

According to different studies in the literature, the most important factors in determining the medical service quality, both from the standpoint of patients and hospital managers, are the issues related to the service provider's human resources,[3,9] which is due to the high amounts of direct interactions or contacts between the employees and customers.[10-12] Therefore, many researchers introduce the service capacity issues as the main resource of the service quality reduction in service-providing organizations such as medical services.[8,13-16] Because of the importance of the service capacity of human resources in human-based service organizations, different studies have been conducted taking "functional quality" into account to determine the desired service capacity based on quantitative methods. However, the level and type of the usage of hospitals from quality improvement differ from one another.[17] Primarily, the analytical models were used by many of the researchers (Table 1).

It is obvious that in analytical models many simplifying assumptions are required for the model to be applicable, and the analyzer is only able to calculate a limited number of system performance criteria in the form of simplified equations. Also, one cannot study the behavior of solutions at the time.[29] With respect to the shortcomings of the analytical models, the simulation models are more efficient in the case of highly complex problems, for they are capable of considering more constraints and enable tracking the variables at the time, which leads to more logical solutions. Also, simulation yields the best solution in complex problems with time dynamics importance.[30] Thus, many of the researchers use the simulation approach to solve the problems related to the human resources service capacity (Table 1).

Despite the fact that the discrete event simulation has widely been the method of interest in determining the capacity and minimizing the delays in human resources capacity determination, in this method, the queue system or service backlog is supposed to be static, and by the course of time, the rate of customer arrival and service delivery are supposed to be invariant, which is contrary to the human-based serving systems for which the assumptions are immensely unrealistic. At the time, the medical service delivery speed, service time per service, and the work pressure (serving system under service pressure) change based on the queue length or service. Actually, human resources in service organizations have variable service delivery rate which is considered in the system dynamics simulation approach.[31,32] In fact, system dynamics is the science that aims to analyze the patterns or feedback or interactive behaviors in complex systems. In other words, system dynamics forms the basis of systemic thinking and is a tool for systemic thinkers (researchers). This analytical approach presents the system behavior by the course of time and enables managers, planners, and policymakers to better understand the system behavior.[33] The system dynamics approach by considering the causal relationships between different model variables and their feedbacks in a systemic view to the human resources capacity issue can help justify the death spirals of quality phenomenon and the policies to cope with them.

Taking advantage of the capabilities of the system dynamics approach, the purpose of this study is to manage the human resources capacity to fight the death spirals of quality using the system dynamics simulation approach, so that the quality of service in medical services is improved. The structure of this study is as follows: first, the conceptual model is developed based on the dynamic hypothesis and the causal loop diagram. Then, the relationships between variables are formulated in the form of the stock and flow diagram. Ultimately, the final model is implemented in a hospital clinic as a case study, and the results of applying different human resources capacity policies on the quality of service are analyzed. The desired policies for escaping from the death spirals of quality are identified and proposed.

Methods

In this research, the delivery gap (difference between the standards or the desired specified states for service delivery and the actual performance of the service provider) in the model of Lovelock and Wright[4] is taken into account, which is defined as the difference between the quality goal (management's perceptions of customers' expectations), quality standard (actual features of the service), and actual quality (difference between the service quality features and the actual service delivered) by Senge and Oliva.[6] Conforming to the standard service times and the level of work errors are considered as the service quality criteria. To analyze the

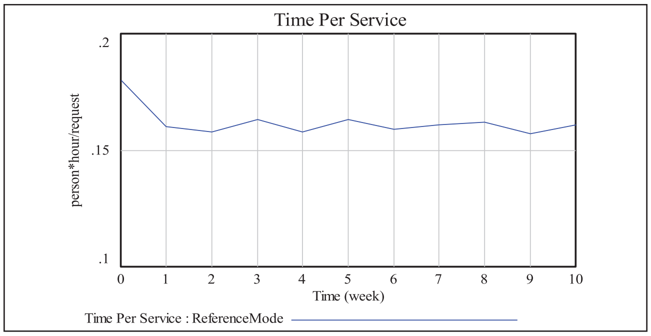

Figure 1. Time per service in case study (reference mode).

death spirals of quality based on human resources dynamics and to appropriately cope with them, we have targeted simulation based on system dynamics approach. System dynamics approach has the purpose of analyzing nonlinear behavior or patterns of complex systems; in other words, it is the basis for systemic thinking. This analytical approach, by simulating the system's behavior over time, helps managers, planners, and policymakers to fully understand the underlying behavior and to examine the results of applying different policies for efficiently controlling and managing this behavior. Using the system dynamics approach, presented in this research seeks to show how the human resources capacity can lead to getting trapped into death spirals of quality and the policies to cope with such situations.

For a better understanding of the designed model, it is implemented in a hospital clinic. The case study here is an eye specialist hospital in Iran (Mashhad city). The physicians of one of the shifts of the clinic are the studied employees. The main issue raised in this clinic was obtained from the behavior of time per service; in other words, according to this clinic's standards, the standard time of each service (the visit time of each patient) is considered about 15 minutes or 0.25 hours; however, as it is shown in Figure 1, the actual time per service is less than the standard time. Although this might seem trivial, with the help of systems dynamics approach, we are willing to prove that this seemingly simple problem has certain complexities. In this case, the system enters a death spirals whose outputs would be the decrease in the time of each service (as the main variable representing service quality). In such situation, policy-making strategies, to solve this problem, will also have particular complexities.

To dynamically study the system, the following 4 main steps were followed in Vensim DSS V6.4 software environment:

Modeling Problem: Dynamic Hypothesis and Causal Loop Diagram

In this research, the dynamic hypothesis of the service quality, concentrating on the human resources service capacity, is finalized after modification and completion based on the literature and hospital managers' opinions, which is as follows:

> "n the medical service system, when the human resources service capacity is lower than its desired level, (based on patients' demand) causes work pressures. In response to work pressures two main reactions are prompted. First, in handling such a situation the employees (physicians) will eventually decrease the service time per service compared to the standard level (cutting the time corners), so that the rate of task completion is increased, which consequently leads into lower service backlog, and as a result, decrease in work pressure (balancing loop). On the other hand, decreasing the service completion time (patients' visit time) results in higher probability of error generation and increase in rework, and ultimately, with the increase in service backlog, the work pressure increases (strengthening loop). The decrease in service time compared to the standard time (due to the lesser contacts between physicians and nurses) and increase in work errors (here error is defined as prescription errors) are the signs of quality erosion in such systems. In order to handle this situation, which is also called as the death spirals of quality, managers would choose to adjust their human resources due to the generated service gap (human resource service capacity being less than its desired level), which leads into more work pressure and quality erosion. To do so, managers increase the number of human resources based on the human resources service gap, by increasing the hire rate (desired hire rate). If the employees (physicians in this research) of a medical service unit are divided into two groups of rookies and experienced employees with different input and output rates and productivity, by the time, the rookies mature into experienced employees and their productivity will increase. In this situation, with the increase in the hired human resources; the quit rate, productivity variations by the time, human resources total productivity, and the number of effective human resources, increase. The more the effective human resources, the more the human resources service capacity; hence, the less the work pressure (balancing loop). On the other hand, hiring new employees (rookies) and quitting of the experienced ones decrease the total productivity. In this situation, the probability of error generation is increased, and with the increase in reworks, the service backlog grows bigger, that leads into more work pressure (strengthening loop).

The mentioned relationships in the dynamic hypothesis are shown in a more detailed view in the causal loop diagram as in Figure 2. The arrows represent the relationships between the variables. A positive (+) or negative (–) sign is assigned to causal relationships, which represent the polarity. A positive relationship means that the changes in the cause and effect are both on the same direction, whereas a negative relationship shows otherwise.

Model Formulation: Stock and Flow Diagram

In this step, the relationships between the system components are formulated from the conceptual (qualitative) form to the quantitative form using the appropriate mathematical

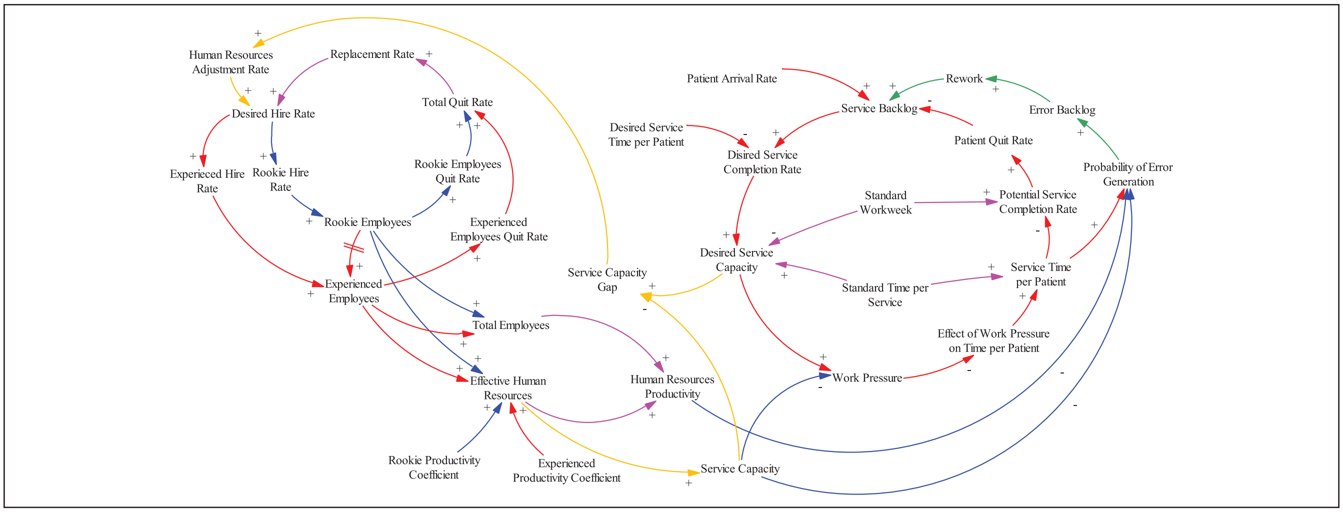

Figure 2. Causal loop diagram.

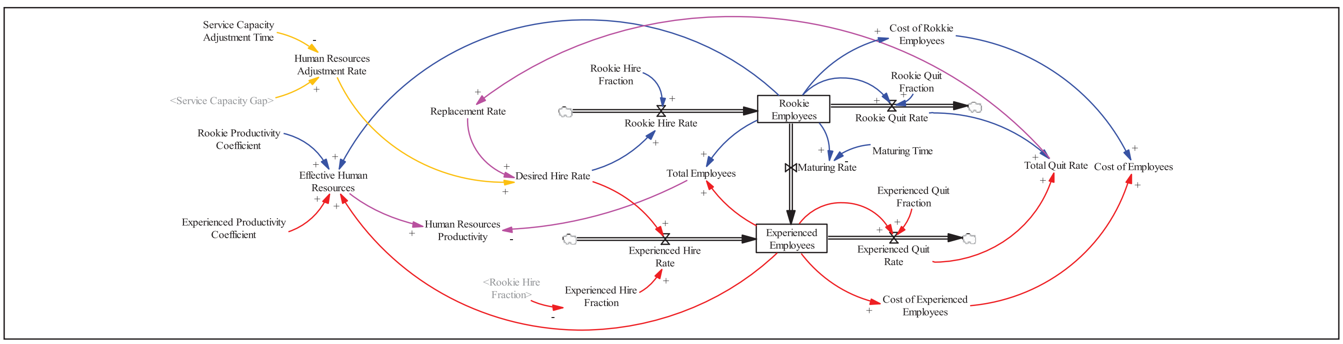

Figure 3. Stock and flow diagram (human resources subsystem).

equations. The stock and flow diagram of the under study system is presented in 2 subsystems of human resources and service capacity.

Stock and flow diagram: Human resources subsystem. The first part of the stock and flow diagram consists of the human resources subsystem (Figure 3). The human resources can be divided into rookies and experienced employees, for which the total number of employees is the summation of them, as in Equation 1 (formulas are presented in Appendix).

Rookie employees is defined as a level variable, the input (hire) rate and output (quit) rate of which change based on Equation 2. In this equation, integral operators are used. The number of rookie employees increases conforming to the rookie hire rate (Equation 3). Also, the number of rookies decreases conforming to the rookies' quit rate (Equation 5). On the contrary, after a specific time period (maturing time), rookies become experienced (Equation 4). Therefore, maturing rate works as a quit rate variable for rookies. The same equations hold for experienced employees (Equations 6-8), only the maturing rate acts as an input variable for experienced employees' stock variable. In other words, when the rookies mature, the number of experienced employees increases. The total employee quit rate can be calculated from the summation of rookies and experienced employees' quit rates (Equation 9). Because in a service delivery system the quitted employees must be replaced, the replacement variable is defined, which is equal to the total quit rate and is considered as the desired hire rate. Obviously, the productivity of the rookies and experienced employees differs due to the difference in the experience level. In other words, the productivity of a rookie employee is less than an experienced one. Thus, a productivity coefficient is considered for each of the experienced and rookie employees. According to Equation 10, multiplying the number of employees by the productivity rates yields the effective human resources variable. Finally, the human resources productivity can be calculated from the effective human resources divided by the total employees, as shown in Equation 11.

Stock and flow diagram: Service capacity and service quality subsystem. The second part of the stock and flow diagram is dedicated to the service capacity and service quality subsystem, which is presented in Figure 4. Generally, cutting the time

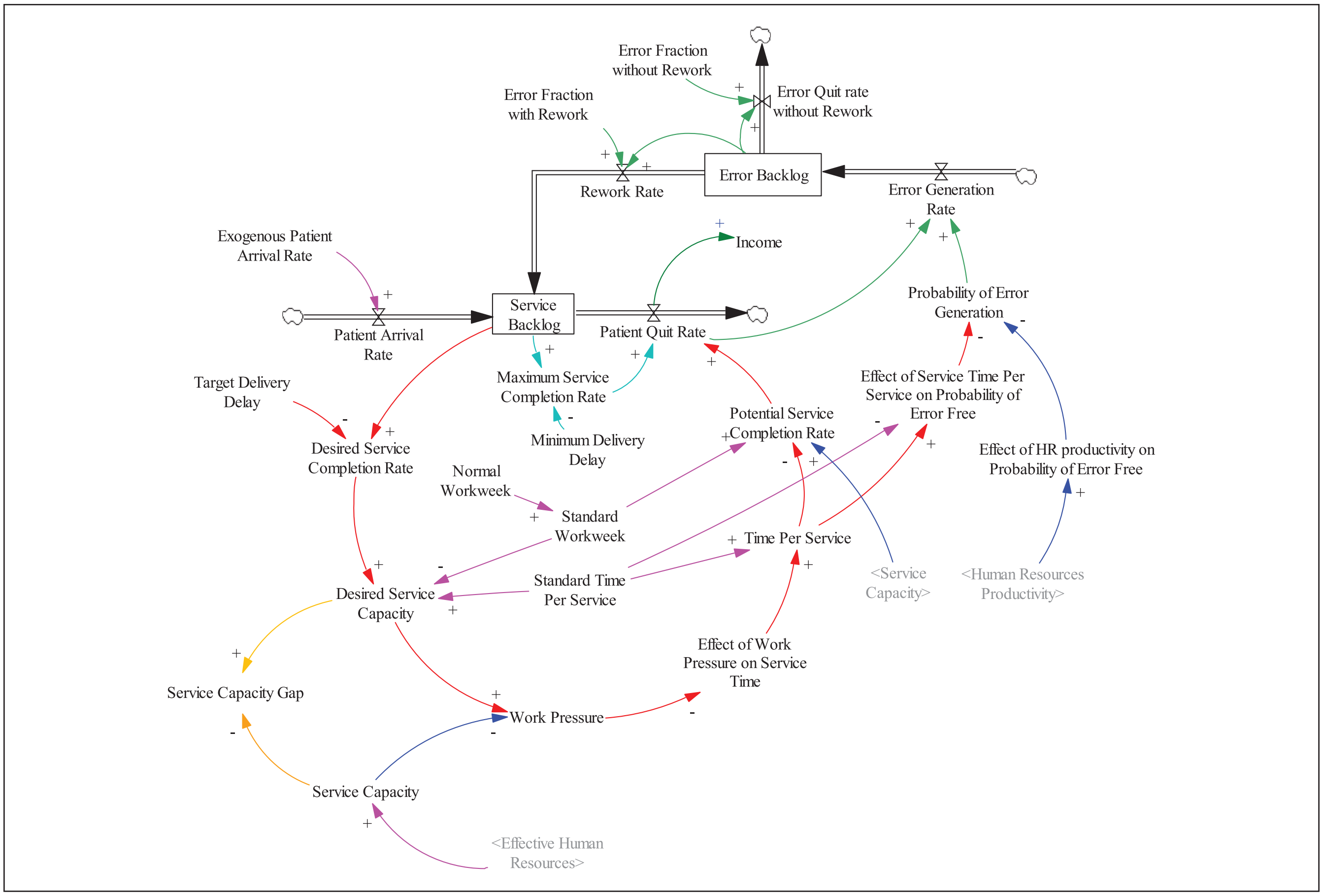

Figure 4. Stock and flow diagram (service capacity subsystem).
Note. HR = human resources.

corners (ie, decreasing the service times compared with the standard times) along with the level of work errors is supposed to be the main reason of falling into the death spirals of quality in this research. In this subsystem, service capacity is equivalent to the number of effective human resources, which is also defined as the intersection of the 2 subsystems (Equation 13). On the other side of the subsystem is the desired service capacity, which can be determined given the desired service completion rate, standard time per service, and the standard workweek, as shown in Equation 14. The desired service completion rate is calculated from the service backlog divided by the desired delivery delay (managers' desired target service time; Equation 15). Service backlog is a level variable, for which the patients' arrival and quit rates act as its input and output rates, respectively (Equation 16). Should a difference exist between the human resources service capacity and the desired service capacity, the system is exposed to work pressure. The work pressure auxiliary variable is considered to be the quotient of the desired service capacity over the actual service capacity (Equation 17). Therefore, values greater than 1 represent the shortage of actual service capacity compared with the desired service capacity, whereas the values less than 1 represent the surplus of the actual service capacity compared with the desired service capacity, such that a portion of the employees is unnecessary in the service delivery system. Finally, the values equal to 1 represent the equality of the desired and actual service capacities. Work pressure causes the employees to decrease the service time per service (cutting the time corners) to respond to the existing demands. At this state, based on Equation 18, high work pressure (greater than 1) leads to a service time less than the standard time to handle the task. The impact of work pressure on the service time can be determined as a function of work pressure estimation. As mentioned before, cutting the time corners (delivery gap = standard time per service – time per service) is an obvious proof of service quality erosion. The potential service completion rate demonstrates the human resource potential in service delivery and can be calculated via Equation 19. As the potential service completion rate cannot exceed the maximum service completion rate (which is determined based on the lowest service delivery time), patients' quit rate is considered as the minimum of the maximum service completion rate and the potential service completion rate, as shown in Equation 20, for which the maximum service completion rate can be calculated from the service backlog divided by the minimum delivery delay (Equation 21). Rework rate is one of the discussed

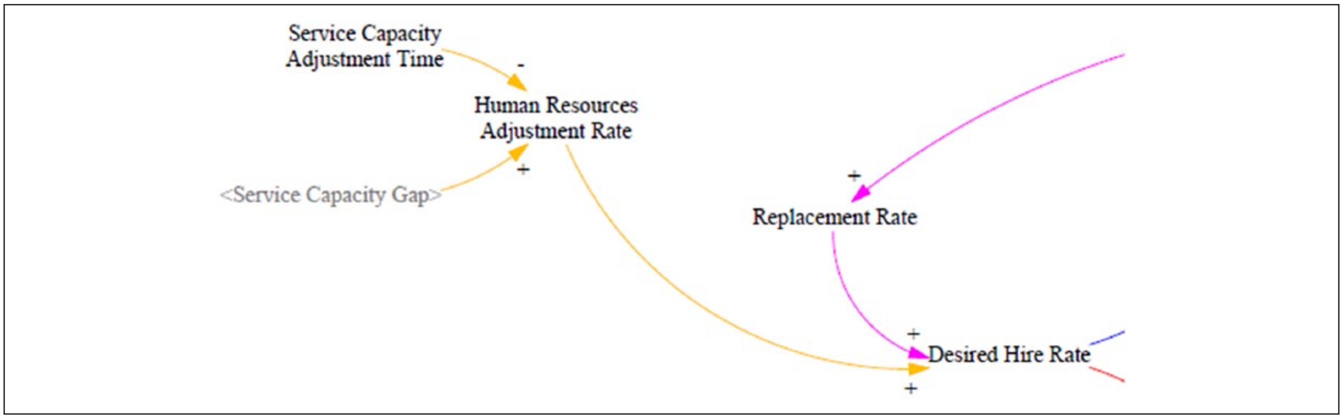

Figure 5. Connection between 2 subsystems.

variables in service backlog. Reworks are due to the errors in performing tasks. Nevertheless, it must be mentioned that not all of the actual errors are identifiable and some of them might even need no reworks. Hence, if the service backlog variable is defined as the error generation rate as the input and error quit rate with rework (Equation 24) and error quit rate without rework (Equation 24) as the outputs, the error quit rate with rework must be added to the service backlog (Equation 22). As noted before, the level of generated errors is a suitable index for service quality erosion. In Equation 24, the error generation rate can be obtained from the product of the patient quit rate and the probability of error generation (Equation 25), in which the probability of error generation can be calculated from Equation 26.

In other words, one of the most important error causes is the decrease in the service time per service compared with the predetermined standard (Equation 27). On the contrary, based on Equation 28, the decrease in human resources productivity is another cause of error generation. Human resources adjustment is the connection between the 2 subsystems of the stock and flow diagram.

Connection between human resources subsystem and service capacity subsystem is delineated in Figure 5. As can be seen, the existence of the service gap requires the service delivery system to adjust its human resources. After the gap identification, as the human resources adjustment is a time-consuming process, a variable representing the service capacity adjustment is embedded in the model. Therefore, the desired hire rate can be obtained from the sum of the replacement rate and human resources adjustment rate (Equation 29).

In the above equation, the human resources adjustment rate equals the service capacity gap divided by the service capacity adjustment time.

Validation Test

After the model is designed, its structure is approved by the experts' opinions. The model behavior is also evaluated using the sensitivity analysis test, which focuses on the reactions (behavior) of the model when exposed to change in parameter values. This test is generally conducted through analyzing the sensitivity of the model to different parameter values. By changing the values of different parameters, the logical behavior of the model is analyzed and validated.

Simulation and Policy Evaluation

After the designed model is validated, different policies for service capacity improvement, focusing on the quality of service delivery, are designed and evaluated. In policy evaluation stage, the effects and consequences of each policy are scrutinized based on the "if-then" analysis, existing sensitivities (system sensitivity analysis), and the behavioral patterns emerging from the simulation of the key variables that represent the service quality improvement (including work errors, cutting the corners of the time per task), and after reviewing the effects of implementing all of the policies, the desired policy is selected. Based on the discussed steps, the results of simulation in the case study are presented in the following section.

Results

In the case study, the time span is supposed to be 48-week time period of 2017. Therefore, the data of patient arrival and quit rates for that time period were obtained from the hospital database. The clinic employees comprised 6 physicians (4 rookies and 2 experienced physicians). The standard time per visit is supposed to be 15 minutes or 0.25 hours. Total weekly working hours for a physician is 40 hours per week (normal weekly working time); the target delivery delay (service time for 1200 patients in the first level of service backlog) and the lowest delay are supposed to be 1 week, each. Due to clinical limitations, it was not possible for the morning shifts to have extra hours of work; therefore, it was not discussed in the model.

According to the hospital managers, the productivity of rookies is less than the experienced physicians, so the productivity rate for rookies and experienced physicians is supposed to be 0.6 and 1, respectively. A 24-week time period is

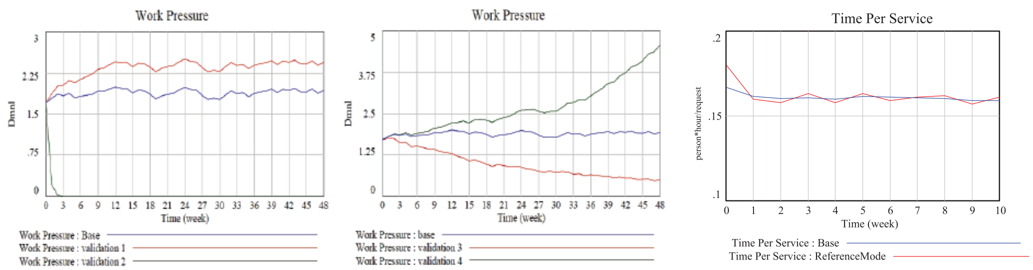

Figure 6. Results of validation tests.

considered as the maturing time. As the rookie physicians do not usually quit before the maturing time in the studied system, the rookies' quit rate is considered 0. On the contrary, based on the historical data, usually 3 experienced physicians quit the system after 3 weeks. Hence, the experienced physicians, quit rate is considered to be 0.1 on average.

For validation test, the sensitivity of the work pressure to the patient's arrival rate and human resources service capacity is demonstrated in Figure 6. In the left diagram, the behavior of the model when arrival rate is multiplied by 1.05 (validation 1) and 0 (validation 2) is compared. As expected, the work pressure increased with the increase in patient arrival rate, and when the patient arrival rate was multiplied by 0, the work pressure before the third week of the time period reached the value of 0. In the right diagram, the sensitivity of the work pressure to the human resources service capacity is shown. The results of multiplying the human resources replacement rate by 2 (validation 3) and 0.9 (validation 4) are compared with the original values. As can be seen, increase in the human resources replacement rate results in the increase in the hire rate, which decreases the work pressure. Decrease in the human resources replacement rate, on the contrary, increases the work pressure.

Moreover, using historical data, the model was validated. For this purpose, the time per service was studied in 2 modes of simulation of the model and reference mode (historical data related to the time per service) over a period of 10 weeks. The results indicated that the simulated values were close to the actual values of time per service.

The next steps after validation are simulation and model behavior analysis based on different policies on a real-world case study, which are provided in the following sections.

The simulation of the designed model is carried out in 3 modes. The first mode consists of analyzing the system behavior while the current values of the exogenous variables are maintained (base mode). Regarding that the current mode is the situation of being trapped in the death spirals of quality, 2 policies are proposed for escaping from the death spirals of quality, which comprise the next 2 simulation modes. The "decrease in patient arrival rate" and "increase in human resources" policies are taken into account in the second and third simulation modes, respectively. The impacts of these 3 simulation modes on the key variables in system behavior analysis including service backlog, service capacity gap (difference between the desired and actual service capacities), work pressure, service time, and the probability of error generation are compared in Figure 7. In the following, the 3 modes of simulation along with their impact on system behavior are further discussed in detail.

No Change in the Exogenous Variables or the Current Mode (Base Mode)

If the variables of demand (number of patients arriving to the system) and human resources (the number and combination of human resources) are considered as the main exogenous variables of the system, due to many different reasons such as imposed macro policies, cost, and income issues, it might be impossible to change them. Therefore, the value of the exogenous variables is supposed to be constant, which is a representation of the current situation, in some ways. As it is obvious in Figure 7, the results of simulating the base mode show that the service capacity gap is always greater than 3, which is the sign of a lack of consistency between the current service capacity and the existing service demand. In such a situation, a heightened work pressure is normally expected, for which the results of the simulation prove so. In this mode, the work pressure value is greater than 1.7, which leads to cutting the service time corners. Hence, the service times become less than the standard time of 0.25 hour (15 minutes) and are in the 0.15- to 0.16-hour interval (9-9.6 minutes). In other words, the physicians due to the imposed work pressure, instead of spending 15 minutes per patient in each visit, decrease the service time to 9 minutes and this confirms the presence of death spirals. Meanwhile, the error generation probability is between 0.2 and 0.3. These are the signs of service quality erosion that are generated in causal and feedback relationships in a dynamic state, also called as the death spirals of quality.

To cope with the death spirals of quality, 2 general policies can be considered. The first policy is decreasing the patient arrival rate and the second policy is increasing the human resources, both of which act to increase the desired

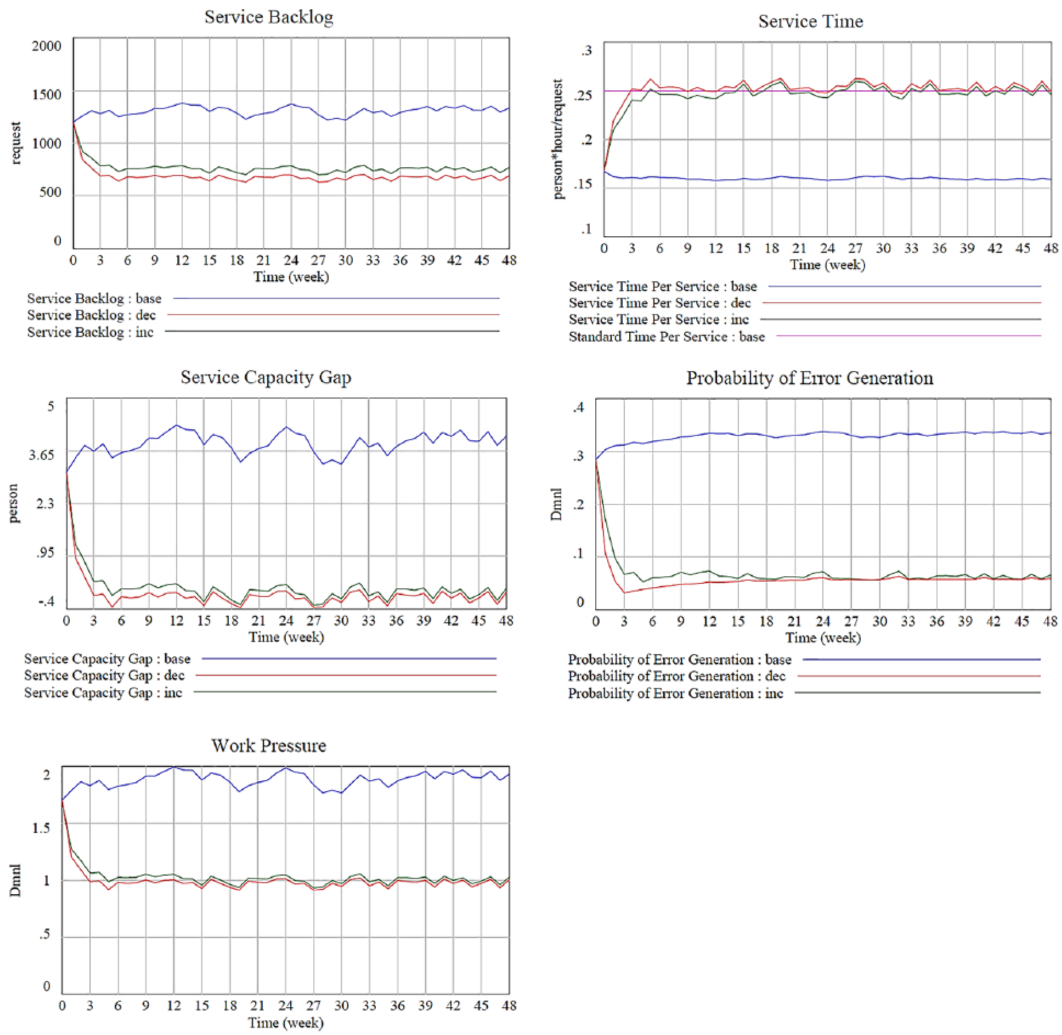

Figure 7. Comparison of simulation results in 3 modes.

service capacity, which results in lower work pressure. These policies are discussed in the following sections.

Policy 1—Decrease in Patient Arrival Rate (Dec)

Decreasing the patient arrival rate (ie, limiting the patients check-in), decreases the service backlog, which leads to lower desired service capacity and eventually lower work pressure. The magnitude of reduction, such that the work pressure becomes suitable, can be estimated using trial and error, numerous variations in patient arrival rate, and analyzing the simulation results. Changing the patient arrival rate finally revealed that a 10% decrease in patient arrival rate yields the desirable results. As shown in Figure 7, applying this policy significantly decreases the service backlog, and then the service capacity gap almost gets to 0. From the third week on, the work pressure approximately vanishes, whereas the visit time per patient is about 0.25 hours (15 minutes). On the contrary, the error generation probability drops drastically. Nevertheless, it must be mentioned that due to the human resource productivity issues, some of the work errors are inevitable.

One must consider that the current conditions of the system might impede the implementation of such policies. Also, decreasing the patient check-in by 10% might not be in full conformity with the general policies and hence be inapplicable. Therefore, the system needs to come up with another policy to compensate for the shortage in service capacity and service quality improvement, which increases the number of human resources. In the third mode, the results of this policy are scrutinized.

Policy 2—Increase in the Human Resources (Inc)

The more the human resources, the more the service capacity. However, it must be noted that the extensive increase in the number of human resources can, on one hand, increase the

Figure 8. Comparison of simulation results in 3 modes (cost of hiring and income).

unemployment rate of the human resources, and, on the other hand, impose unnecessary costs to the system. Therefore, it is crucial for the human resources to be increased such that the desired service capacity is not violated and the work pressure is neutralized. Trial and error method unveiled that a 30% increase in the human resources level can yield the desired results. Based on this finding, if the number of the physicians increases from 6 to 8 (with the same assumption of the new employees to be rookies), the results are pretty similar to the ones of the policy 1, that is, a service capacity gap of 0 and a work pressure fluctuating about 1. Also, implementing this policy causes the service (visit) time to reach 0.25 hours, that is, hiring 2 additional physicians can increase the service capacity such that the physicians are able to spend the standard amount of time (15 minutes) per patient. Also, due to the smaller difference between the actual and standard service times, the error generation probability drops drastically.

The application cost and the incomes obtained from the patients' visits are shown in Figure 8. Studies have shown that the first policy (10% reduction in the patient's entry rate) would result in an income loss of 2 660 000 Rials (Iran's currency) per week compared with the base mode (lost opportunity cost). And the use of the second policy (a 30% increase in the number of physicians) will also cost an average of 1 509 000 Rials compared with the base mode. Because the goal is to reduce the difference between time per service and standard time service, the second policy seems more effective.

In the case study, it is assumed that all of the newly hired physicians are rookies and their quit rate is 0. Obviously, any change in the hire rate of the rookies and experienced physicians, and their quit rates, leads up to different system behaviors.

As previously mentioned, the system needs to hire human resources to prevent from the quality erosion resulting from the insufficient human resources. In the proposed policies, the hire rate is supposed to be 0. In this section, the hire rate variable (with a 3-week time to mature) is also included in the model to enable the dynamic human resource hiring for the system. The simulation results are reported in Figure 9. It is evident that with the increase in the human resources level, the human resources service capacity gap has almost vanished from about the 10th week forward. Hence, the work pressure fluctuates with a negligible difference about 1, which is the sign of a pressureless serving system. It is expected that cutting of time corners is decreased until it is totally disappeared. According to the results, from the 10th week on, the time standards (0.25 hours) are applicable in the serving system. Also, the error generation probability is significantly decreased. Hence, regarding the simulation results, it can be concluded that the performance of the simulation model in improving the service quality concentrating on the human resources is desirable.

Limitations

The present research studies the doctors of a single hospital clinic, while a chain of serving clinics were active at the time. The human resources service capacity of each clinic affects the managing approach service capacity of the studied clinic, which is not taken into account in this research due to many limitations such as time and lack of data.

Discussion

Correct human resource management and planning in medical service delivery systems can immensely prepare the foundation of quality improvement. Having a systemic thought that considers the complexities and dynamics of such systems can assist the hospital managers in building a profound knowledge of the system and deal with the human resources problems more realistically. In this regard, this study presented a dynamic model to show the cycles of causal and feedback relationships that eventuate in quality erosion, also called as the death spirals of quality. The designed model consisted of 2 subsystems of human resources service capacity and the causal relationships between the key variables, along with the connection between the 2 subsystems. Finally, the model was simulated for the physicians of a real-case hospital clinic. The simulation results point that regarding the number

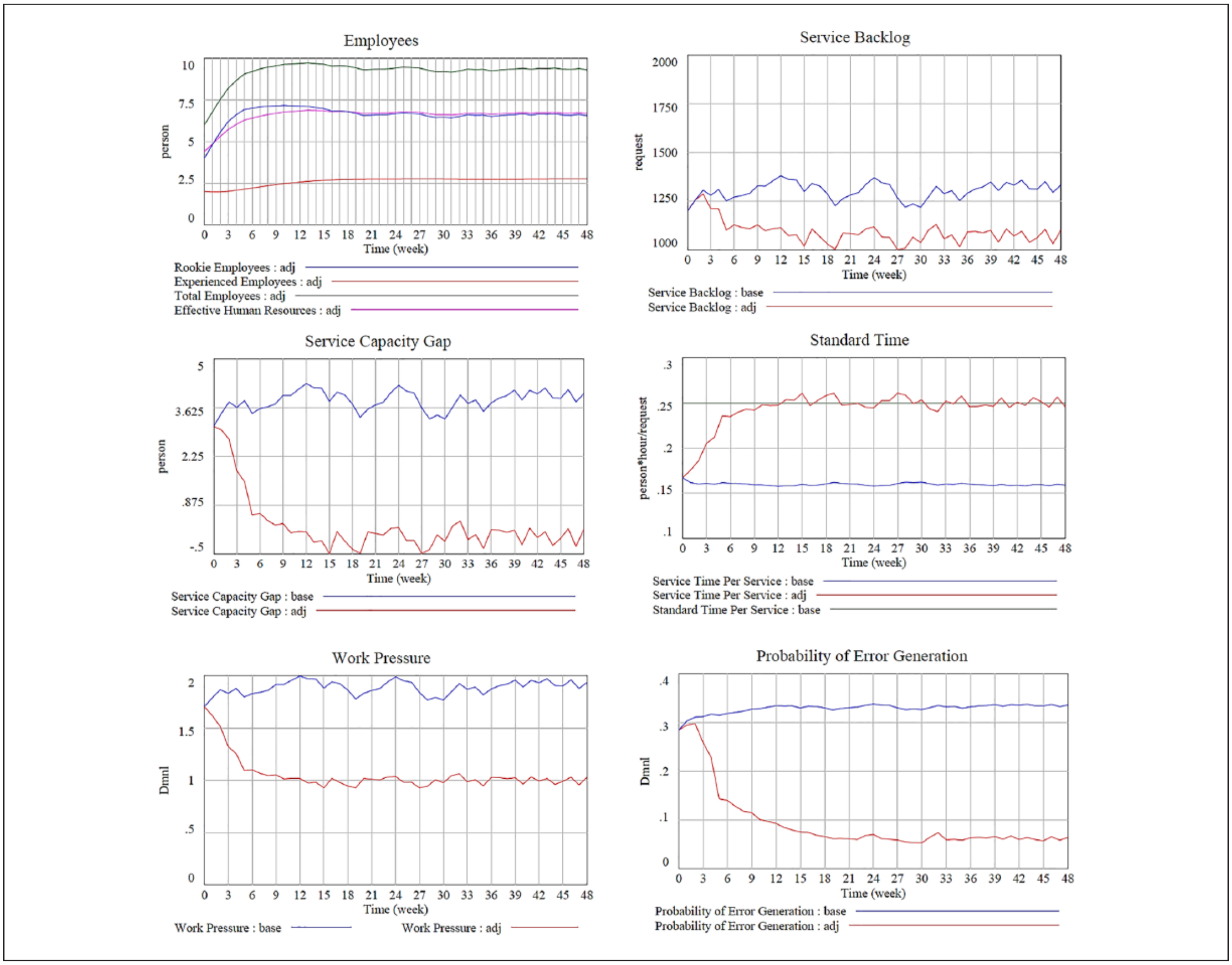

Figure 9. Simulation results in human resources adjustment.

of patients in the system, the existing human resources service capacity was incapable of responding to all of the service backlogs. This situation forces the employees to decrease the service time per service due to the imposed work pressure, which is one of the most important problems that arise because of the lower linguistic contacts between the service deliverer and the patient in medical service systems. This issue, by itself, increases the patients' dissatisfaction and quality erosion. Also, the decrease in service time increases the error generation probability that can, on one hand, impose reworks, and, on the other hand, make a fertile field for patient dissatisfaction. To deal with such situations, which is the true symptom of the death spirals of quality, 2 general policies were proposed based on the simulation results. The first policy is limiting the patients arriving to the system (commensurate with the conditions of the studied system) to decrease the service backlog and balance the work pressures. However, applying this policy may not always be possible due to the macro policies and existing constraints. Therefore, the second policy is increasing the human resources (commensurate with the conditions of the studied system) to correctly answer the service backlog, to improve the quality.

In complex modeling of human resource planning, both analytical models (such as linear programming) and simulation models (such as system dynamics) have been used; however, it should be noted that for complex issues in which dynamics of time is of importance, simulating models are the best choice. On the contrary, in other simulation approaches (such as discrete event approach), the system has been statically viewed and the speed of customer arrivals and service delivery are assumed fixed in the system. However, these assumptions can be relatively unrealistic in the service systems in which the service provider is human. Because over time, depending on the length of the queue and the work pressures (service under pressure system), the service providers change the speed of service delivery and the time they spend for each customer. In fact, in service organizations, human resources do not have a stable service offering over a period of time, and therefore this has been specifically considered in human resource planning based on the dynamics of the system in the present study. All in all, it can be said that using systems dynamics approach in managing human resources service capacity of the human-centered service systems would provide more realistic analyses compared with other approaches.

Conclusions

The simulation results revealed that determining the required number of human resources in a dynamic system can extensively prepare the system to apply the standard service time and reduce work pressures. On the contrary, human resources hiring at the time in a dynamic framework was made possible, such that the number of hired human resources, to make the system apply the standard service times and reduce work errors, was dynamically determined. Implementing the designed model helps the hospital managers to act systemically in dealing with human resource planning for quality improvement. Albeit, it must be noted that increasing the number of human resources in a system that has been working under work pressure for a while might not yield the desired results. For instance, when the human resources reduce the service time based on the current service conditions, after a while, they get used to the service times, such that even increasing the number of human resources might not reset the service time per service at the standard level. To not get trapped in such situations, the hospital managers need to justify the human resources and by applying the supervisory systems prepare for performance improvement.

Implications

In the present research, the appropriate policies to deal with quality death spirals in medical services (focusing on conforming to the time standards and work errors) are studied with a dynamic approach. The number and experience of the human resources are considered as the human resources productivity criteria. In future research works, it can be suggested that the impacts of the human resources education on the productivity be taken into account. On the contrary, regarding that in the selected case study the overtime and its impact on service capacity is not included, including this issue can be of great importance in future research works. Also, the problem was modeled for only a hospital unit. The death spirals of quality can be studied as a series of dependent units considering the relationships between them in the future studies.

Appendix

Formulas and Details in Stock and Flow Diagram.

a) Human resources subsystem:

Total Employees = Rookie Employees + Experienced Employees (1)

Rookie Employees = INTEG (Rookie Hire Rate – Maturing Rate – Rookie Quit Rate, Initial Stock) (2)

Rookie Hire Rate = Desired Hire Rate × Rookie Hire Fraction = Replacement Rate × Rookie Hire Fraction (3)

Maturing Rate = Rookie Employees/Maturing Time (4)

Rookie Quit Rate = Rookie Employees × Rookie Quit Fraction (5)

Experienced Employees = INTEG (Experienced Hire Rate + Maturing Rate – Experienced Quit, Initial Stock) (6)

Experienced Hire Rate = Desired Hire Rate × Experienced Hire Fraction = Replacement Rate × (1 – Rookie Hire Fraction) (7)

Experienced Quit Rate = Experienced Employees × Experienced Quit Fraction) (8)

Total Quit Rate = Rookie Quit Rate + Experienced Quit Rate = Replacement Rate = Desired Hire Rate (9)

Effective Human Resources = (Experienced Employees × Experienced Productivity Coefficient)+(Rookie Employees × Rookie Productivity Coefficient) (10)

Human Resources Productivity = Effective Human Resources/Total Employees (11)

Cost of Employees= Cost of Rookie Employees + Cost of Experienced Employees = (Rookie Employees × 700 Rials) +(Experienced Employees × 900 Rials) (12)

Details: The human resources can be divided into rookies and experienced employees, for which the total number of employees is the summation of them, as in Equation 1 (Rookie employees is defined as a level variable, the input (hire) rate and output (quit) rate of which change based on Equation 2. In this equation, integral operators are used. The number of rookie employees increases conforming to the rookie hire rate (Equation 3). Also, the number of rookies decreases conforming to the rookies' quit rate (Equation 5). On the contrary, after a specific time period (maturing time), rookies become experienced (Equation 4). Therefore, maturing rate works as a quit rate variable for rookies. The same equations hold for experienced employees (Equations 6-8); only the maturing rate acts as an input variable for experienced employees' stock variable. In other words, when the rookies mature, the number of experienced employees increases. The total employee quit rate can be calculated from the summation of rookies and experienced employees' quit rates (Equation 9). Because in a service delivery system the quitted employees must be replaced, the replacement variable is defined, which is equal to the total quit rate and is considered as the desired hire rate. Obviously, the productivity of the rookies and experienced employees differs due to the difference in the experience level. In other words, the productivity of a rookie employee is less than an experienced one. Thus, a productivity coefficient is considered for each of the experienced and rookie employees. According to Equation 10, multiplying the number of employees by the productivity rates yields the effective human resources variable. Finally, the human resources productivity can be calculated from the effective human resources divided by the total employees, as shown in Equation 11. The cost of employees is calculated in Equation 12.

(continued)

Appendix. (continued)

b) Service capacity and service quality subsystem:

Service Capacity = Effective Human Resources (13)
Desired Service Capacity = (Desired Service Completion Rate × Standard Time Per Service)/Standard Workweek (14)
Desired Service Completion Rate = Service Backlog/Desired Delivery Delay (15)
Service Backlog = INTEG (Patient Arrival Rate + Rework Rate – Patient Quit Rate, Initial Stock) (16)
Work Pressure = Desired Service Capacity/Service Capacity (17)
Service Time Per Service = Standard Time Per Service × Effect of Work Pressure on Service Time = Standard Time Per Service × f (Work Pressure) (18)
Potential Service Completion Rate = (Service Capacity/Service Time Per Service) × Standard Workweek (19)
Patient Quit Rate = Min (Maximum Service Completion Rate, Potential Service Completion Rate) (20)
Maximum Service Completion Rate = Service Backlog/Minimum Delivery Delay (21)
Error Backlog = INTEG (Error Generation Rate-Error Quit rate without Rework-Rework Rate, Initial Value) (22)
Error Quit rate without Rework = Error Backlog × Error Fraction without Rework (23)
Rework Rate = Error Backlog × Error Fraction with Rework (24)
Error Generation Rate = Patient Quit Rate × Probability of Error Generation (25)
Probability of Error Generation = 1 – Probability of Error Free = 1-1 – (Effect of Service Time Per Service on Probability of Error Free × Effect of Human Resources productivity on Probability of Error Free) (26)
Effect of Service Time Per Service on Probability of Error Free = f(Service Time Per Service) (27)
Effect of Human Resources productivity on Probability of Error Free = f(Human Resources productivity) (28)
Desired Hire Rate = Replacement Rate + Human Resources Adjusted Rate = Replacement Rate + (Service Capacity Gap/ Service Capacity Adjusted Time (29)
Income = (Patient Quit Rate × 25 Rials) (30)

***Details*:** In this subsystem, service capacity is equivalent to the number of effective human resources, which is also defined as the intersection of the 2 subsystems (Equation 13). On the other side of the subsystem is the desired service capacity, which can be determined given the desired service completion rate, standard time per service, and the standard workweek; as shown in Equation 14. The desired service completion rate is calculated from the service backlog divided by the desired delivery delay (managers' desired target service time; Equation 15). Service backlog is a level variable, for which the patients' arrival and quit rates act as its input and output rates, respectively (Equation 16). Should a difference exist between the human resources service capacity and the desired service capacity, the system is exposed to work pressure. The work pressure auxiliary variable is considered to be the quotient of the desired service capacity over the actual service capacity (Equation 17). Therefore, values greater than 1 represent the shortage of actual service capacity compared with the desired service capacity; whereas the values less than 1 represent the surplus of the actual service capacity compared with the desired service capacity, such that a portion of the employees is unnecessary in the service delivery system. Finally, the values equal to 1 represent the equality of the desired and actual service capacities. Work pressure causes the employees to decrease the service time per service (cutting the time corners) to respond to the existing demands. At this state, based on Equation 18, high work pressure (greater than 1) leads to a service time less than the standard time to handle the task. The impact of work pressure on the service time can be determined as a function of work pressure estimation. As mentioned before, cutting the time corners is an obvious proof of service quality erosion. The potential service completion rate demonstrates the human resource potential in service delivery and can be calculated via Equation 19. As the potential service completion rate cannot exceed the maximum service completion rate (which is determined based on the lowest service delivery time), patients' quit rate is considered as the minimum of the maximum service completion rate and the potential service completion rate, as shown in Equation 20, for which the maximum service completion rate can be calculated from the service backlog divided by the minimum delivery delay (Equation 21). Rework rate is one of the discussed variables in service backlog. Reworks are due to the errors in performing tasks. Nevertheless, it must be mentioned that not all of the actual errors are identifiable and some of them might even need no reworks. Hence, if the service backlog variable is defined as the error generation rate as the input and error quit rate with rework (Equation 24) and error quit rate without rework (Equation 24) as the outputs, the error quit rate with rework must be added to the service backlog (Equation 22). As noted before, the level of generated errors is a suitable index for service quality erosion. In Equation 24, the error generation rate can be obtained from the product of the patient quit rate and the probability of error generation (Equation 25), in which the probability of error generation can be calculated from Equation 26. The income is calculated in Equation 30.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Alireza Pooya https://orcid.org/0000-0001-6000-3535

References

1. Sohail MS. Service quality in hospitals: more favorable than you might think. *Manag Serv Qual*. 2003;13(3):197-206.

2. Hollis C. *Service Evaluation in Internal Healthcare Service Chains* [master's thesis], Brisbane, Australia: Faculty of Business, Queensland University of Technology; 2006.
3. Chaker M, Al-Azzab N. Patient satisfaction in Qatar Orthopedic and Sports Medicine Hospital (ASPITAR). *Int J Bus Soc Sci*. 2011;2(7):69-78.
4. Lovelock CH, Wright LK. *Principles of Service Marketing and Management*. 2nd ed. London, England: Prentice Hall; 2002.
5. American Customer Satisfaction Index (ACSI). http://www.theacsi.org. Accessed July 18, 2016.
6. Senge PM, Oliva R. *Developing a theory of service quality/service capacity interaction*. The Conference on Transforming Organizations, Sloan School of Management, MIT; May 29-31, 1993; Cambridge, MA.
7. Shaikh U, Romano P, Paterniti DA. Organizing for quality improvement in health care: an example from childhood obesity prevention. *Qual Manag Health Care*. 2015;24(3):121-128.
8. Oliva R, Sterman JD. Death spirals and virtuous cycles: human resource dynamics in knowledge-based services. In: Maglio P, Spohrer J, Kieliszewski C, eds. *Handbook of Service Science*. New York, NY: Springer; 2010:321-358.
9. Handayani PW, Hidayanto AN, Sandhyaduhita PI, Kasiyah A, Ayuningtyas D. Strategic hospital services quality analysis in Indonesia. *Expert Syst Appl*. 2015;42(6):3067-3078.
10. Zeithaml VA, Parasuraman A, Berry LL. Problems and strategy in service marketing. *J Marketing*. 1985;49(2):33-46.
11. Sasaki H. The rise of service employment and its impact on aggregate productivity growth. *Struct Change Econ D*. 2007;18(4):438-459.
12. Lupo T. A fuzzy framework to evaluate service quality in the healthcare industry: an empirical case of public hospital service evaluation in Sicily. *Appl Soft Comput*. 2016;40(1):468-478.
13. Oliva R, Sterman JD. Cutting corners and working overtime: quality erosion in the service industry. *Manage Sci*. 2001;47(7):894-914.
14. Linderman K, Schroeder RG, Choo AS. Six Sigma: the role of goals in improvement teams. *J Oper Manag*. 2006;24(6):779-802.
15. Oliva R, Bean M. Developing operational understanding of service quality through a simulation environment. *Int J Serv Ind Manag*. 2008;19(2):160-175.
16. Bendoly E. Real-time feedback and booking behavior in the hospitality industry: moderating the balance between imperfect judgment and imperfect prescription. *J Oper Manag*. 2013;31(1-2):62-71.
17. Mora AM, Walker D. Quality improvement strategies in accountable care organization hospitals. *Qual Manag Health Care*. 2016;25(1):8-12.
18. Willemain TR, Moore GT. Planning a medical practice using paramedical personnel. *Health Serv Res*. 1994;9:53-61.
19. Smith KR, Over AM, Hansen MF, Golladay FL, Davenport EJ. Analytic framework and measurement strategy for investigating optimal staffing in medical practice. *Oper Res*. 1976;24(5):815-841.
20. Kwak NK, Lee C. A linear goal programming model for human resource allocation in a health-care organization. *J Med Syst*. 1997;21(3):129-140.
21. Jaumard B, Semet F, Vovor T. A generalized linear programming model for nurse scheduling. *Eur J Oper Res*. 1998;107(1):1-18.
22. Mobasher A. *Nurse Scheduling Optimization in a General Clinic and an Operating Suite* [master's thesis], Houston, TX: University of Houston; 2011.
23. Lin CC, Kang JR, Liu WY. A mathematical model for nurse scheduling with different preference ranks. *Future Inf Technol*. 2015;329:11-17.
24. Punnakitikashem P, Rosenberger JM, Behan DB. Stochastic programming for nurse assignment. *Comput Optim Appl*. 2008;40(3):321-349.
25. Jayasinha Y. Decreasing turnaround time and increasing patient satisfaction in a safety net hospital–based pediatrics clinic using Lean Six Sigma methodologies. *Qual Manag Health Care*. 2016;25(1):38-43.
26. Rohleder TR, Lewkonia P, Bischak DP, Duffy P, Hendijani R. Using simulation modeling to improve patient flow at an outpatient orthopedic clinic. *Health Care Manag Sci*. 2011;14(2):135-145.
27. Ghanes K, Jiuini O, Jemai Z, Wargon M, Hellmann R, Thomas V, Koole G. *A comprehensive simulation modeling of an emergency department: a case study for simulation optimization of staffing levels*. Proceedings of the Winter Simulation Conference; December 7-10, 2014; Savannah, GA.
28. Gul M, Fuat Guneri A. *A discrete event simulation model of an emergency department network for earthquake conditions*. Proceedings of the 6th International Conference of Modeling Simulation and Applied Optimization, American University of Sharjah; May 27-29, 2015; Sharjah, United Arab Emirates.
29. Banks J, Carson JS, Nelson BL, Nicol DM. *Discrete-Event System Simulation*. 4th ed. New York, USA: Prentice Hall; 2005.
30. Borshchev A, Filippov A. *From system dynamics and discrete event to practical agent based modeling: reasons, techniques, tools*. The 22nd International Conference of the System Dynamics Society; July 25-29, 2004; Oxford, England.
31. Acker AV, Haxholdt C, Larsen ER. Long-term and short-term customer reaction: a two-stage queuing approach. *Syst Dynam Rev*. 2006;22(4):349-369.
32. Lan CH, Chang CC, Kuo MP. Service system under service pressure by system dynamics model. *ProbStat Forum*. 2010;3(1):52-64.
33. Kumar D, Kumar D. Modelling rural healthcare supply chain in India using system dynamics. *Procedia Engineer*. 2014;97:2204-2212.
34. Evans GW, Gor TB, Unger E. *A simulation model for evaluating personnel schedules in a hospital emergency department*. Proceedings of the 28th Conference on Winter Simulation, IEEE Computer Society; December 8-11, 1996; Washington, DC.
35. Duguay C, Chetouane F. Modeling and improving emergency department systems using discrete event simulation. *Simulation*. 2007;83(4):311-320.
36. Al-Araidah O, Boran A, Wahsheh A. Reducing delay in healthcare delivery at outpatient clinics using discrete event simulation. *Int J Simul Model*. 2012;11(4):185-195.