Queueing Theory – A Tool for Production Planning in Health Care

Adeniran S. Afolalu, Oluwasegun J. Adelakun, Samson O.Ongbali, Abiodun A. Abioye, Oluseyi O. Ajayi, Oloyede Olamilekan R

Abstract- This paper provides an assessment of recent modeling of processes and mathematical approach in operations research which is applicable to the analysis of waiting there by enabling individuals to estimate the number of resources required to meet a particular demand in hospital. The main essence of modeling using mathematical methods is to derive formulae that allow system to evaluate system performance using an equation. The study examined four key areas, making queuing model difficult some of which include; Changes in preventive interruption, re-appearing of patients and vague channeling, Modeling of interruption and capacity utilization and focus on models that can be used in hospital with limited and unlimited rooms. We concluded that the best model is the unlimited waiting room models which limitation have higher utilization that can attend to patients at any available time. In order to improve the performance of effectiveness and productivity in healthcare, there is need to discover and apply relevant models.

Index Term --- Health care, Hospitals, models, Production planning, Queuing theory

I. INTRODUCTION

Queuing theory is a modeling and mathematical approach in operations research that is applied to waiting lines, thereby enabling individuals to estimate the resources necessary to meet the needs [1]. Various attempts have been carried out by applying queuing analysis into hospital activities by splitting them into various departments as a means of ensuring adequate delivery [2].

Queues was analyzed in the context of telephone facilities, so as to ensure reduction in waiting time; planning and extra investments is required [3]. Background study in stochastic analysis and probability hypothesis is required in studying queuing so as to understand its functionality in real situation, especially in hospital where competence is required to respond to demands regarding the patient mix and unevenness in the arrival flow in order to minimize the delay in any health care process is done by queuing models [4]. Understanding its functionality in real situation, especially in hospital where competence is required to respond to demands regarding the patient mix and unevenness in the arrival flow in order to minimize the delay in any health care process is done by queuing models [4]. Health care capacity planning involves assuming the quantity of resources necessary to deliver health care services at particular cost and quality [5]. Increase in population of health-need has resulted in increase in waiting times and overcrowding in hospital especially in Emergency Departments [6]. Healthcare professionals discovered the importance of applying queuing theory techniques in which many have been discouraged by the mathematical mystery regarding patients and staffs [7].

This study summarize a range of queuing theory in determining the optimal capacity level that minimizes cost while maintaining a desired level of performance on patients' service quality and financial boundaries. The main essence of modeling using mathematical methods is to derive a formula that allows system to evaluate system performance using an equation [1]. The aim is to provide awareness to everyone interested in locating and modeling queuing theory for health sector process by making the models, reasonable to understand, adapt, and apply [6].

II. LITERATURE REVIEW- CHALLENGES

The patient has to be delayed for a long period of time before being attended to because of delay resulting from waiting time, moreover the capacity needs of some emergency departments cannot match the patients attendance and time of waiting [1]. Difficulty in understanding the models is another serious challenge with the principle of aggregation in handling doubt with less capacity [8]. Increase in patients attendance makes it difficult for the patient to wait for their turn once they see the queue is too long, they prefer to rather turn back or abandoned to other health care having a lesser queue [9].

Manuscript received March 23, 2019; revised April 02, 2019. This work was supported in part by Covenant University. Ota. Nigeria..
Adeniran S. Afolalu is a Lecturer from Mechanical Engineering Department. Covenant University Ota. Nigeria. Sunday.afolalu@covenantuniversity.edu.ng +2348035785758
Oluwasegun J. Adelakun, Samson O.Ongbali, Abiodun A. Abioye, Oluseyi O. Ajayi are Lecturers from Mechanical Engineering Department Covenant University Ota. Nigeria.
Oloyede, Olamilekan R is a Lecturer from Mechanical and Mechatronics Engineering Afe Bablola University Ado-Ekiti. Nigeria.

Overcrowded waiting rooms with patients as a result of inadequate bed space, inability of managing and allocating surgical theaters among the specialists [10]. When a patient joins a queue and wait until it get to his/her turn before enter the service station to be attended to, the queuing systems may benefit patients conduct in the form of defections from the queue in which a patient may change to the queues that is moving quickly[11]. Reneging is also achievable in the queue whereby a patient wait on queue for some time and then exit the system as a result of being a gradual process; therefore patient reneges probability increases with the queue length and in a system where the number of patients exceeds the capacity of the server making reneging the only possible route for that system to attain balance equilibrium [9]. .Some patients on the other hand, may decide not to join the queue for some reason and may decide to return for the service later and this situation is known as balking [12].

2.1 Changes in preventive interruption

As a result of the service interruption in healthcare been a major source of concern, consensus on the negative effects of interruption exist in patient flow times including the quality of service with evidence of healthcare environments characterized by considerable amounts of variability which induces waiting times [9]. Preventive interruption whenever a physician is disrupted by an impromptu emergency or calls when attending to patients makes models that is been developed to include disruption that is certain to occur during service time [13]. After proper examination of patients by the physician, patients may be told to undergo some tests to validate physician action resulting in patients moving from one department to another [14].

Movement of patients to one or more department in hospital is not determined in many instances where lots of examination has to be done before moving the patient to the theatre for Surgery and after surgery; monitoring examination is needed to ensure the patient is recuperating leading him/her to keep appearing at different times thereby making it difficult to use a queuing model as a result of the ambiguousness created on new patients [8]. Majority of queuing models allocate different events like surgery, consultation of patients and nursing mothers at different time and consider a particular period of time to be constant [15]. Equipments and materials are not available in most hospitals and if available in some hospitals; it will not be able to meet up with the number of patients to be served thereby resulting in the division of time in areas needing surgery and aspect requiring consultation [16]. Hospitals operation requires high capacity utilization so as to have a large impact on its performance due to the number of patients arriving at every time, which has made materials to be inadequate, leading to network stations to be congested with crowd of patients requiring urgent attention [17]. Situations resulting from emergency, staffs not available, disruption during service occurring in hospital make it impossible to know the exact time a hospital operates or the time the doctor will be available for consultation [14]. Unplanned changes are what leads to increase in the number of patients waiting for service; because more time is required to attend to the patients, developing a model for overtime is complex [25].

2.2 Health Care Modeling System

The Poisson process is used to model the arrival process for queuing [21]. The Poisson process is a continuous-time process. N(t) is the number of arrivals in [0, t] for a Poisson process with rate, λ, such that the time between successive arrivals is exponentially distributed with parameter [21]. Analytically, a process is said to be Poisson if patients arrive differently from one another, thereby making its postulation of different arrivals the most commonly used arrival processes in modeling service systems [20]. The Poisson process has been confirmed to be accurate for impromptu arrivals of patients to various parts of the hospital. The limitation of this process is that it can be characterized by a single parameter which is the rate [13]. The pure birth-death process is used to examine the number of patients in a queue. It is a case of an uninterrupted stochastic process which are of two types: "deaths" which decrease the condition by lone and births" which increase the condition by lone [22].

The Birth process is arrival process while death process is departure process. The two processes describes the number of patients probabilistically, N (t) in the queuing system changes with time (t) increases [23]. The limitation of this model is that the two processes are mutually independent. The M/M/1 model is made up of a server that render service to patients who arrive and depart from the system peradventure a data from a patient arrives at the time the server is full of activity, then the patients have to join the waiting line [24]. Two types of events occur and they include; arrival events and departure events which are used in representation of the model [25]. This model may consider for modeling patient's arrival, drugs to be administered, patient examination and provide services where the arrival of the patient changes with time. Capacity requirements are estimated when the number of patients present in the system can not affect the disturbance of patient arrival time [22]. One benefit of the M/M/s model is that only three parameters are required and performance estimates with minimal data can be obtained. The number of servers, s, easy-to-compute formulae are available to obtain performance measures such as the probability that an arrival will experience a positive delay or the average delay [26]. The limitation of this model is the high usage of its server due to overcrowding of waiting patients. Moreover, using the model may not be realistic when the cost of patient waiting is higher, having excess capacity is also not ideal when the cost of materials is larger than the cost of serving the patients with the materials [24]. In addition the M/M/1 Model has a single server serving patients on a first come where the population is very large, so that arriving patients are unaffected by the size of the queue. The model's delay waiting times due to server unavailability [26]. Assume an increase in the amount of time for which a patient has to be served is not advisable .Because instances of service times in some laboratory services may not vary much from one patient to another; it may be more accurate in some instances to model using a method of distribution instead of increasing thereby making it suitable for mathematical calculating [24]. Sharing of number in the system and patients delay can be calculated numerically using present inventions in the aspect of change in numerical conversion [27]. However, the limitation of this model is that the expressions for distribution are bound are not easy to derive [24].

This model can have an unlimited number of servers in which each of the servers is instantly provided for each patient arriving in the system [28]. This model makes necessary minimum belief concerning the nature of arrival that exist within and time of service that need to be shared among three models of the unlimited waiting room. Analysis of this model necessitates solving Lindley's integral equation [22]. Particular approaches fall into two categories which includes; approximating either of the constant of this model and assuming distribution that is rigid before estimating its parameters. A lot of theories have been developed for estimating the average number the system in this model [27]. The study of this model enlightens service providers in limiting changes in time of arrival within and allocating the waiting time available by the day on the website so that patients can program their time of arrivals to the hospital when the doctors and member of staffs are not busy [24].

2.3 Limited Waiting Room Models

Congestion in hospitals and its various departments is to a certain extent common because when rooms becomes occupied, arriving patients or patients on queue leave without getting been served, In order to model this situation, models having patients available in service and the total number of patients being reduced must be considered [29]. If the limit at which patients can be admitted in the hospital is exceeded, outcome of occurrence can either be modeled normally by discouraging patients from joining the queue pending the period that waiting room is vacant or allowing patients to keep adding to the queue but withdraw immediately the waiting room is occupied [25]. The emergency department can make use of this model to further enhance its capacity. The approach to the model can be modified to include a book showing that the expected patient time of consultation.[26] .

The M/G/1/K depend upon the finishing point from embedded Markov chain which has resulted to different equations which can be numerically analyzed to making the outcome of occurrence to add up to one in order to derive a stable method of making number in the system evenly distributed [27]. The limitation of this model is that if the probability does not sum up to 1 then there can never be a stable state of making number in the system being distributed [28]. While The GI/G/1/K model is not easy to compute or calculate except in cases where the arrival of patients within the hospital with the time of service being allocated to each patients have in features in common various publications whose topics is on the computation of this model has postulated different theories showing the relationship between the limited waiting room and the unlimited waiting room models [29]. This model is a more precise in which a limit of K patients is permitted in the system; when the system contains K patients, the arriving patients are rejected since there can never be more than K patients in the system, the system reaches a balanced state when it has attended to the number of patients that it can serve for the period [30]. This model can also be called the M/M/C loss system because patients who enter as at the time all servers are active are not permitted to wait to be served but rather exit [31]. The possible outcome of occurrence stating all servers are active so that an arriving patient leaves the system, is called "Erlang's B formula" the limitation of this model is the issue of no waiting for patients [32].

2.3.1 Servers in a Single Station Model

This model having unlimited waiting line constitute major part of the system and the most universally studied models of unlimited waiting room have gained more popularity in study with two out the three unlimited waiting rooms models mathematically tractable, which in turn comes from M in the model descriptor signifying the presence of Markovian with assurance stating the arrival of patients within the hospital is been trigger by the increase in sharing patients to a service time in majority of health care [33]. .If $\rho > 1$ explains that for every time the server becomes available, issues from arriving patients is greater than unity [24]. In situations where $\rho = 1$, patients waiting might increase as a result of the uncertainty in different arrivals and time of service; it leads to idleness of server because for every issue an average amount of patients brings at every time the server is available is exactly unity [35]. The consequence of not being busy makes the patients waiting to increase, leading to stable distribution. If $\rho < 1$ in all models with unlimited waiting rooms then there is desire to indicate the service pattern so as to evaluate the queue the limitation of this model is the relationship dependency between arrivals leading to the inability of Poisson probability distribution to be strictly applied when the input population is finite [27].

This describes the facilities available in a hospital with different servers handling patients waiting in line for service. In such instances, the time at which service is provided to its patients is a factor of patient number on the system with the overall server utilization given as $\rho = \lambda \; m\mu$, where the focus is on one of the unlimited waiting rooms models in which different arrival of patients and the time allocated for service are increasing and distributed among the patients to be served [19]. Expressions of the mean delay in this model can be effectively put to use in deriving vital output in the design of waiting line system that puts together the number of patients waiting and the server that minimizes delay from the system [36]. Queuing network models have been examined and there are lots of applications of the models. The model requires adequate techniques to obtain system performance and measures [20]. A wide variety of healthcare system makes use of the treatment materials at their disposal in performing services for a medical situation, making it to be best modeled as connection of waiting patients [38].

Lack of capacity in reaching out to those at the grass root handling facilities with adequate and necessary information might make patient, to be delayed for a long period of time for specified days [37]. This model makes use of different methods of approach in explaining path patients follow so as to get an appointment using the instance of getting appointment in the health care system. [33].

2. 4 Priority Queue

Doctors in Emergency Department attend to patients whose condition are extremely critical, so as to ensure their survival. [64]. In some hospitals, doctors make certain slots available as a means of reserve for patients whose state of condition and respect are highly considered Moreover, when a patient whose state of condition and respect is not recognized by the hospital, he/she will have to wait till it gets to his/her turn [65]. In some instances, absentee at work as at when due result in priority rules, adding to the time it will take the service provider will be available to patients across various unit of the health care.[66]. There are two priority principle which includes the static and dynamic. In a static priority principle, every single patient has an appointment, but other patients arriving to the same waiting room is given appointment over the all the rest of patients having appointment, but in the case of dynamic priority principle, patients that have been given an appointment higher than another at a particular department may be given an appointment lower than the previous [39]. For example, when a doctor gives an appointment to Mr John as patient number two out of thirty in the hospital waiting, if the doctor decided to tell each as they are examined to go for the test, when they all returned, Mr. John might be on number twenty for appointed to see the doctor again out of thirty and that is what dynamic priority principle is stating. In addition, doctors give high priority to patients whose cost of waiting with time can be attended to within a short period [26].

2.4.1 Queueing theory a tool in healthcare.

In most hospitals, especially in Ota, it is common to find patients waiting for diagnostic test or surgery, thereby leading to delay; in another to decrease this delay which has led to the poor satisfaction of patients, financial implication; the focus of many hospitals within and outside Ota is to make queuing analysis extremely valuable in utilizing resources so as to minimize delay. By optimizing entry of patients across each department in the hospitals so as to provide solution to average patient wait time, average visit time, probability of patient waiting beyond half an hour and below; length of average patient has been occupied by doctor, probability of hospital referring patients to another hospital and progress on patient entry at every point [40-42].

III CONCLUSIONS

Reviewed has looked into the essence of different healthcare processes. The survey reviewed that limited waiting room models are not easy to analyze and no waiting time for patients while , unlimited waiting rooms models have high server utilization and a bound expression of distribution. Single server in a single station has a dependency between arrivals, multiple servers in a multiple station making it difficult to obtain bound expression. Network model sometimes has inadequate capacity at the grass root facility and priority queue in health care require higher waiting cost for quick response. We concluded that the best model is the unlimited waiting room models which limitation has higher utilization that can attend to patients at any available time. In order to improve the performance of effectiveness and productivity in healthcare, there is need to discover and apply relevant models. However, there is a necessity of putting into consideration different methods on how the waiting line in a health care system interact by formulating models for future use

ACKNOWLEDGEMENTS
We acknowledge the financial support offered by Covenant University in actualization of this research work for publication.

REFERENCE

[1] Jolanta,l(2015).Improving the production planning and control process, 13(4).
[2] Palvannan, R. K., & Teow, K. L. (2012). Queueing for healthcare. Journal of Medical Systems, 36(2), 541–
[3] Wang, D. (2016). Delay prediction with enhanced queuing models and survival analysis.2016 3rd International Conference on Systems and Informatics (ICSAI).
[4] . Lin, C. (2008). Optimization Models for Capacity Planning in Health Care. Journal of Operational Research, 29(5), 36–46.
[5] Olorunsola, S. A., Adeleke, R. A., & Ogunlade, T. O. (2014). Queueing Analysis of Patient Flow in Hospital. IOSR Journal of Mathematics (IOSR- JM), 10(4), 47–53.
[6] Giambene, G. (2014). Survey on Probability Theory. In Queuing Theory and Telecommunications (pp. 265–317).Computer Science, 19(1), 1–18.
[7] Bastani, P. (2009). A Queueing Model of Hospital Congestion. Simon Fraser University, 1–74.
[8] Fomundam, S., Herrmann, J., Fomundam, S. F., & Herrmann, J. W. (2007). The Institute for systems research a survey of Queuing Theory applications in healthcare A SURVEY OF QUEUING THEORY APPLICATIONS IN HEALTHCARE.
[9] Fomundam, S., Herrmann, J., Fomundam, S. F., & Herrmann, J. W. (2007). The Institute for systems researches a survey of Queuing Theory applications in healthcare A SURVEY OF QUEUING THEORY APPLICATIONS IN HEALTHCARE.
[10] Shanmugam, R. (2014). How do queuing concepts and tools help to efficiently manage hospitals when the patients are impatient ? A demonstration, 2(3), 1076–1084.
[11] Rodríguez Jáuregui, G. R., González Pérez, A. K., Hernández González, S., & Hernández Ripalda, M. D. (2017). Análisis del servicio de urgencias aplicando teoría de líneas de espera. Contaduria y Administracion, 62(3), 733–745.
[12] Mba, J. C. M., & Noon, C. (2011). The Definitive Guide to Emergency Department Operational Improvement: Employing Lean Principles with Current ED Best Practices to Create the “No Wait” Department. CRC Press
[13] Creemers, Stefan, and Marc Lambrecht. 2007. “Modeling a Healthcare System as a Queuing Network: The Case of a Belgian Hospital.” SSRN Electronic Journal.
[14] Salawu, E.Y., Okokpujie, I.P., Afolalu, S.A.,Ajayi, O.O. and Azeta, J., 2018. INVESTIGATION OF PRODUCTION OUTPUT FORIMPROVEMENT. International Journal of Mechanical and Production Engineering Research and Development, 8(1), pp.915-922
[15] Afolalu, Sunday A. and Ayuba, Samuel U. and Ihebom Ikechi V. and Elewa, Remilekun R. and Kehinde (2018) ROLE OF RELIABILITY MANAGEMENT TOOL ANDDOCUMENTATIONS- REVIEW. Global Journal of Engineering Science and Research Management, 5 (3). p. 39. ISSN 2349- 4506
[16] Blocker, R. C., Heaton, H. A., Forsyth, K. L., Hawthorne, H. J., El-Sherif, N., Bellolio, M. F., Hallbeck, M. S. (2017). Physician Interrupted Workflow Interruptions and Patient Care in the Emergency Department. The Journal of Emergency Medicine, 53(6), 798–804
[17] Salawu, E. Y., Ajayi, O. O., Inegbenebor, A. O., Afolalu, S. A., & Ongbali, S. O. (2018). PARETO ANALYSIS OF PRODUCT QUALITY FAILURES AND COST EFFECTS IN BOTTLINGMACHINES- LEAN THINKING SOLUTION FOR ALCOHOL INDUSTRY. International Journal of Mechanical Engineering and Technology, 9(11), 2380-2388.
[18] Maurya, V. N. (2010). On the expected busy beriod of an interdependent M/M/1:(∞;Gd) queuing model using bivariate Poisson process and controllable arrival rates. In 2010 2nd International Conference on Electronic Computer Technology https://doi.org/
[19] Afolalu, A. S., Enesi, Y. S., Kehinde, O., Samuel, U. A., Ikechi, V. I., & Remilekun, R. E. (2018) Failure Mode and Effect Analysis a Tool for Reliability Evaluation. European Journal of Engineering Research and Science, 3(4), 65-68.
[20] Dushime, A., Mung, J. K., & Ndengo, M. (2015). Queuing Model for Healthcare Services in Public Health Facilities (A Case Study of Muhima Hospital), 3(1), 77–93.
[21] . Akkerman, R., & van Donk, D. P. (2008). Analyzing scheduling in the food-processing industry Structure and tasks. Cognition, Technology & Work, 11(3), 215–226.
[22] . Buestán, M. (2015). Determining the most appropriate Production Planning and Control system for Small Enterprises : framework and field tests, 1–195.
[23] . Cai, Y., Kutanoglu, E., & Hasenbein, J. (2011). Production Planning and Scheduling : Interaction and Coordination.
[24] Gilaninia, S. (2016). Optimizing Production Planning and Environmental Assessment of Conformity with the Inventory Management Methods in Iran Optimizing Production Planning and Environmental Assessment of Conformity with the Inventory Management Methods in Iran, (December 2011).
[25] . Graves, S. C. (2014). A Review of Production Scheduling, (August 1981).
[26] Jolanta,l(2015).Improving the production planning and control process, 13(4).
[27] Ivy, J. S., Huschka, T. R., Rohleder, T. R., & Marmor,

Y. N. (2013). CAPACITY Management and Patient Scheduling in an Outpatient Clinic Using Discrete Event Simulation, (2005), 2215

[28] Jinsoo Park, Park, J., Lee, H., So, B., Kim, Y., Kim, B. H., … Park, B. C. (2014). New key performance indices for complex manufacturing scheduling. In Proceedings of the Winter Simulation Conference 2014

[29] . Kallrath, J. (2003). Planning and scheduling in the process industry, 1–36.

[30] Afolalu, A. S., Enesi, Y. S., Kehinde, O., Samuel, U. A., Ikechi, V. I., & Remilekun, R. E. (2018). Failure Mode and Effect Analysis a Tool for Reliability Evaluation. European Journal of Engineering Research and Science, 3(4), 65-68.

[31] Moniz, S., Barbosa-póvoa, A. P., & Sousa, P. De. (2015). On the complexity of production planning and scheduling in the pharmaceutical industry : the Delivery Trade-offs Matrix, (June).

[32] Reschke, J. (2017). Scheduling parameters in Production Planning and Control, 15(1), 62–66.

[33] Schilling, D. A. (2000). Master production scheduling in capacitated sequence-dependent process industries, (May).

[34] Technology, C., Akkerman, R., & Donk, D. P. Van. (2014). Analysing scheduling in the food-processing industry : Structure and tasks, (April).

[35] Volume-, I., & Ngonadi, I. V. (2017). Deploying the Queueing Model in the Remote Patient Medical Monitoring System, (6), 111–118.

[36] Green, L. (2006). Chapter 10 QUEUEING ANALYSIS IN HEALTHCARE. Patient Flow: Reducing Delay in Healthcare Delivery International Series in Operations Research & Management Science, 2006, Volume 91, 281-307

[37] Sobieraj, J. (2015). Queuing Analysis of a General Medical Practice, (May) pg 22

[38] Rotich, T. (2016). Utility Analysis of an Emergency Medical Service Model Using Queuing Theory. British Journal of Mathematics & Computer Science, 19(1), 1–18.

[39] Choudhury, G., & Deka, M. (2012). A single server queueing system with two phases of service subject to server breakdown and Bernoulli vacation. Applied Mathematical Modelling, 36(12

[40] Mardiah, F. P., & Basri, M. H. (2013). The Analysis of Appointment System t o Reduce Outpatient Waiting Time at Indonesia's Public Hospital. Human Resource Management Research, 3(1), 27–33.

[41] Yu, M., Tang, Y., Fu, Y., & Pan, L. (2011). An M/Ek/1 queueing system with no damage service interruptions. Mathematical and Computer Modelling, 54(5–6), 1262–1272.

[42] Gilaninia, S. (2016). Optimizing Production Planning and Environmental Assessment of Conformity with the Inventory Management Methods in Iran Optimizing Production Planning and Environmental Assessment of Conformity with the Inventory Management Methods in Iran, (December 2011).

ISBN: 978-988-14048-6-2
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)