Measuring effective capacity in an emergency department

Björn Lantz
Technology Management and Economics, Chalmers University of Technology, Gothenburg, Sweden, and
Peter Rosén
Department of Business Administration, University of Gothenburg, Gothenburg, Sweden

Received 1 February 2014
Revised 10 May 2014
22 August 2014
17 September 2014
Accepted 18 September 2014

Abstract

Purpose – The purpose of this paper is to show how elements from queueing theory can be used to obtain objective measures of effective capacity in the triage function at Skaraborg Hospital in Sweden without direct observation of the function itself.

Design/methodology/approach – Approximately 30,000 patients arrived to the emergency department at Skaraborg Hospital in Sweden during 2011. The exact time of arrival and the exact time of triage were recorded for each patient on an individual level. Basic queueing theory uses arrival rates and system capacity measures to derive average queueing times. The authors use the theoretical relation between these three measures to derive system capacity measures based on observed arrival rates and observed average queueing times.

Findings – The effective capacity in the triage process is not a linear function of the number of nurses. However, the management of capacity seems well adapted to the actual demand, even though service levels vary substantially during the day and night.

Originality/value – This paper uses elements from queueing theory in an innovative way to measure the effective capacity of a service process without direct observation, thereby also avoiding the potential risk of the Hawthorne effect.

Keywords Performance measurement, Performance management, Emergency department, Sweden, Performance monitoring, Queueing theory, Triage function, Hawthorne effect, Capacity measurement

Paper type Research paper

1. Introduction

The concept of capacity is central to operations management. For example, in production planning, evaluations of changes in processes and in productivity measurements, it is crucial to have a reliable methodology for measuring the available capacity of the operation. This is especially true in the service sector (Sasser, 1976; Fitzsimmons and Fitzsimmons, 2008), and thus also in the control of health care organizations (Bamford and Chatziaslan, 2009).

Capacity is traditionally defined as the maximum potential output of a process during a specific period, for example, the number of cars manufactured per week or the number of patients treated per day. In this regard, capacity has nothing to do with actual output. Capacity can also be defined functionally in terms of design and effective capacity. Design capacity refers to the maximum output of a process under ideal conditions, that is, the output a process is designed to achieve at full utilization of productive time. Effective capacity is the maximum output of a process under normal conditions. Design capacity is always greater than effective capacity, and the difference

The authors wish to thank Mattias Sundén for mathematical assistance.

Journal of Health Organization and Management
Vol. 30 No. 1, 2016
pp. 73-84

1477-7266
DOI 10.1108/JHOM-05-2014-0074

between them can be explained by productive time under normal conditions being used for things such as shifts in production, on-going maintenance, and breaks. Effective capacity in a process can conventionally be calculated as design capacity multiplied by the proportion of the productive time, which is utilized under normal conditions (Russell and Taylor, 2006).

In processes that produce a limited number of standardized products or services, capacity in terms of output is relatively easy to define and measure. It becomes more complicated to measure capacity in terms of output when the mixture of products or services produced in the process becomes more complex. Capacity during a specific period is then dependent on the mixture of products demanded during the period. In such cases, it can be more appropriate to define and measure capacity in terms of input (Slack *et al.*, 2010). Within health care, for example, capacity is commonly measured in terms of the number of hospital beds (Kuntz *et al.*, 2007). However, this definition of capacity can be called into question, since the time dimension is ignored. According to the definition above, the number of hospital beds does not become a capacity measurement until it is combined with, for example, the number of patients treated per day (Schroeder, 1993).

The design capacity of a process under different assumptions can often be calculated without too much trouble, but the true level of the effective capacity can be more difficult to measure. The reason is primarily that it is difficult to know what normal conditions actually are. An alternative to the above standard calculations of effective capacity is to do empirical measurements, through, for example, surveys or direct observations (Finkler *et al.*, 1993; Sittig, 1993; Bratt *et al.*, 1999; Yen *et al.*, 2009). However, the validity of such results is questionable because of the potential risk of the Hawthorne effect (Roethlisberger and Dickson, 1939; Franke and Kaul, 1978; Campbell *et al.*, 1995; Wickström and Bendix, 2000). According to the Hawthorne effect, a process is affected by the fact that it is observed. For example, staff might work either more or less efficiently than they would under normal conditions because they know that their performance is being measured.

The objective of this paper is to develop a methodology based on queueing theory in order to measure the effective capacity in a service process without direct observation of the process itself, and to use the methodology to measure the effective capacity in the triage process of the emergency department at Skaraborg Hospital in Sweden. The methodology can be applied to any type of waiting line system that can be characterized by queueing theory. Waiting line processes with such characteristics are perhaps most frequent in service environments, such as emergency departments, supermarkets, banks, and restaurants, but they are also common in traditional operations management, in areas such as job shop scheduling, machine loading, and maintenance service.

As mentioned earlier, this paper measures the effective capacity in the triage process of the emergency department at Skaraborg Hospital in Sweden. Historically, operations management in hospitals has generally not relied much on queueing theory or other analytic techniques from management science (Laffel and Blumenthal, 1989), despite the fact that researchers have suggested many such approaches over the years. However, during the subsequent decade, the use of queueing theory has increased in health care (see Lakshmi and Sivakumar, 2013, for a review). For example, Green *et al.*, (2006) evaluated the usefulness of queueing models in guiding emergency department provider scheduling decisions. Izady and Worthington (2012) used elements from queueing theory to propose an algorithm that seeks to stabilize the quality of services delivered by all emergency care providers at all times, Koeleman *et al.* (2012) provided a model to manage the personnel-planning problem in care-at-home service facilities in a stochastic setting based on Markov decision theory, Li and Glazebrook (2011) took a Bayesian approach to the uncertainty generated by error prone triage and discussed the design of heuristic policies for scheduling jobs for services to maximize the mean number of jobs served, Bowers (2011) used elements from queueing theory to simulate waiting list management and analyzes its connection to capacity planning, De Bruin *et al.* (2007) showed how queueing theory could be applied to model emergency cardiac in-patient flow, and Mayhew and Smith (2008) used queueing theory to analyze the UK government's 4-hour completion time target in accident and emergency departments.

Different formula expressions that represent key figures can be calculated, depending on the assumptions regarding the context in which a queue occurs (including the distribution of arrivals, the service process distribution, the number of servers, etc.). For example, these key figures include the expected queue length, expected waiting time, and expected number of "customers" in the system. As most standard text books in the field show (e.g. Balakrishnan *et al.*, 2007), such formulas are in most cases based on two parameters, namely, the expected number of arriving "customers" per unit time and the expected effective capacity, expressed as the number of "customers" served per unit time. As a result, a practical measurement problem when formulas for the analysis of queues are applied to a process is that the effective capacity is assumed to be known.

The arrival rate and queue length (or queueing time) are variables that are usually easy to measure objectively and empirically with high validity. In addition, the relationship between arrivals, effective capacity, and queue length can be expressed mathematically by known models. Therefore, it is possible to use empirical data for arrivals and queue lengths associated with a particular service process to derive the effective capacity of the process. In other words, if the average queue length can be expressed mathematically as a function of arrival rate and processing rate, then it is possible to express the average processing rate as a function of arrival rate and the average queue length. Through empirical measurements of these two variables, the processing rate in terms of the effective capacity of the system can be calculated.

The rest of this paper is organized as follows. In the next section, we show how queueing models can be used to estimate capacity. After that, we use actual data from the emergency department at Skaraborg Hospital to measure their effective capacity. Finally, we conclude the paper, with recommendations for future research in this field.

2. Using queueing models to measure capacity

As indicated in the introduction, the idea here is to obtain an estimate of the effective capacity in terms of the maximum number of jobs that, on average, can be processed per period ($\overline{\mu}$). This estimate is a function of the estimated average number of arrivals per period $\left(\overline{\lambda}\right)$ and the estimated average waiting time in the queue $\left(\overline{W}_q\right)$ using the fact from queueing theory that W_q, under certain assumptions regarding the arrival process and the service process, can be written as a function of λ and μ. When estimates are used to sketch such a relation, it is of course only approximate. Since it is possible to rearrange the formula $\overline{W}_q \approx f\left(\overline{\lambda}, \overline{\mu}\right)$ to obtain a formula where $\overline{\mu} \approx f\left(\overline{\lambda}, \overline{W}_q\right)$, we can use $\overline{\lambda}$ and $\overline{W}_q$ to obtain $\overline{\mu}$. As $\overline{\lambda}$ and $\overline{W}_q$ can be estimated independently of the service process, the effective capacity of the service process can be estimated objectively, thus avoiding the Hawthorne effect.

The basis for the theoretical queue modeling formula used here is that arrivals are assumed to follow a Poisson distribution and service times an exponential distribution when the system consists of one server and the queue discipline is first-in, first-out

(FIFO) with no reneging. Because both arrivals and service are Markovian processes (i.e. they are stochastic processes with no memory of the past) and because there is only one server, the model is called M/M/1 using Kendall notation. A vast array of different assumptions regarding the arrival process, the service process, and/or other system characteristics can be modeled (Gross *et al.*, 2008; Mehdi, 2002).

Under the basic M/M/1 queueing model assumptions, a standard textbook in this area will state that the average waiting time in the queue can be defined in terms of λ and μ as:

$$W_q = \frac{\lambda}{\mu(\mu-\lambda)} \tag{1}$$

Using the relationship in (1), the maximum number of jobs that, on average, can be processed per period can be written as a function of the average waiting time and the average number of arrivals:

$$\mu = \lambda/2 + \sqrt{(\lambda/2)^2 + \lambda/W_q} \tag{2}$$

Using the same idea, expressions for $\mu = f(\lambda, W_q)$ can be obtained for other queueing models. As long as there is only one server, as in models such as the M/M/1, the M/D/1, the M/G/1, or the M/Ek/1, the formulas and their derivation are relatively simple. For more than one server, the situation quickly becomes more complex, however. For the M/M/2 model, for example, the average waiting time is defined in terms of λ and μ as:

$$W_q = \frac{\lambda^2}{4\mu^3 - \mu\lambda^2} \tag{3}$$

Solving (3) for μ yields a relatively complicated formula:

$$\mu = \frac{1}{2}\left(\frac{\lambda^2 W_q}{3^{1/3}\left(9\lambda^2 W_q{}^2 + \sqrt{3}\sqrt{27\lambda^4 W_q{}^4 - \lambda^6 W_q{}^6}\right)^{1/3}} + \frac{\left(9\lambda^2 W_q{}^2 + \sqrt{3}\sqrt{27\lambda^4 W_q{}^4 - \lambda^6 W_q{}^6}\right)^{1/3}}{3^{2/3} W_q}\right) \tag{4}$$

For some more complex queueing models, $W_q = f(\lambda, \mu)$ cannot at all be rearranged to $\mu = f(\lambda, W_q)$, but they can often be used to measure capacity in a similar way by using a simple trial-and-error algorithm to find the value of $\overline{\mu}$ that corresponds with the empirically measured values of $\overline{\lambda}$ and $\overline{W}_q$. Such an algorithm would still be in line with the main idea in this paper; that the effective capacity in a service process can be measured without direct observation of the process itself.

When one uses the point estimates $\overline{\lambda}$ and $\overline{W}_q$ to obtain the point estimate $\overline{\mu}$, one would need to consider the combined uncertainty of $\overline{\lambda}$ and $\overline{W}_q$ when creating a confidence interval for the true effective capacity. However, using estimates of two parameters to estimate a third parameter obviously escalates the uncertainty.

Furthermore, because observed waiting times are dependent, it is not possible to derive a simple expression for a confidence interval for the effective capacity based on the observed waiting times. However, as long as the analysis is based on large quantities of data, we know that a confidence interval for the effective capacity would be narrow.

3. The emergency department at Skaraborg Hospital

An emergency department is a medical treatment facility specializing in the acute care of patients who arrive without prior appointment, either by their own means or by ambulance. The emergency department at Skaraborg Hospital operates 24 hours a day and had approximately 45,000 patient visits in 2011, of which approximately 15,000 arrived by ambulance. Arrivals to the emergency department comprised approximately 65 percent of all arrivals at Skaraborg Hospital. The triage process at the emergency department is an initial medical assessment of patients made by nurses in which patients are categorized by urgency of need. In this process the queueing discipline is strictly FIFO. The patient flow at Skaraborg Hospital through the emergency department is illustrated in Figure 1.

The triage is a standardized process that follows a well-documented workflow with clear guidelines, and so, it is reasonable to measure the capacity in terms of output during a specific period.

The emergency department at Skaraborg Hospital is responsible for its own staff planning with respect to assistant nurses, nurses, and secretaries, but not of doctors. Responsibility for staff planning of doctors rests with the specialist departments. Staff planning in the triage process is based on standard estimates of effective capacity. A triage team consists of two nurses during the day, who are expected to serve, on average, 5.5 patients per hour. During the night, the triage is staffed by a single experienced nurse who is expected to serve 3.5 patients per hour on average. A linear relationship between the number of nurses and the number of patients served is assumed in the planning process at the emergency department. In other words, a doubling of capacity during the day or night is assumed to lead to a doubling of the number of patients served.

This case study is based on the empirical results from interviews with management in the emergency department and a data set describing the arrival and queueing times to triage for 30,000 arrivals to the emergency department during 365 days in 2011. We have excluded the 15,000 patients who arrived at the emergency department by ambulance as they are triaged in the ambulance on the way to the emergency department, and so do not queue for triage. Figure 2 shows the average number of arrivals per hour to the emergency department during 2011.

As the number of arrivals is well known by management to vary during a 24-hour period, the number of nurses in the triage also varies according to a schedule. Hence, the expected capacity of the triage process in terms of the service time is not constant. A potential problem when we apply queueing theory to analyze these data is that both the service time itself and the variance in the service time vary over a 24-hour period. During

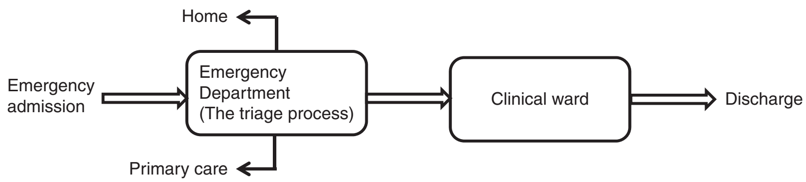

Figure 1. The patient flow through the emergency department at Skaraborg Hospital

the night, the triage is attended by a single, experienced nurse. However, during the day, several nurses work in the triage. When there are two nurses, they work individually, but when there are three or four nurses, they work in two teams. Hence, during the day, there are always two "servers", but there is only one "server" during the night. Thus, the basis for our queue data analysis was M/M/1 during the night, and M/M/2 during the day.

Because arrivals to the ED obviously are cyclic, our overall analytic approach was the well-established Lag stationary independent period-by-period (SIPP) method (Green, 2006; Green *et al.*, 2001). The general idea behind Lag SIPP is that the regular SIPP approach is unreliable in systems with time-varying arrival rates, as the time of peak congestion lags the time of the peak in the arrival rate. For M/M/s queueing models, this lag is well approximated by the average service time (Green, 2011). Thus, our observed arrival rates were advanced by a guesstimation (based on our interviews with ED management) of the average triage time, 15 minutes. Sensitivity analyses where arrival rates were advanced 10 and 20 minutes, respectively, only yielded trivial differences.

First, we needed to ensure that the arrivals could be assumed to be Markovian processes in the individual hours. Since the arrival frequency varies during the day, we divided the day into 24 periods of one hour each and analyzed the Markovian assumption for each hour individually with goodness-of-fit tests. These tests supported the assumption of Poisson distributions in the individual hours. Hence, we proceeded under the assumption that arrivals in each hour do reflect a Markovian process.

The next step was to use the observed average number of arrivals and the observed average waiting time for each hour to estimate the average effective capacity for each hour using the methodology presented in the previous section. Table I presents the relevant estimates.

Figure 3 displays the hourly effective capacity. As expected, the estimated effective capacity is higher during the day, and it seems relatively well adapted to variation in the arrival process.

In Figure 4, the hourly estimated effective capacity is graphed alongside the actual number of nurses. Not surprisingly, there is coherence, albeit far from perfect. At night, the capacity is relatively high, while the capacity during the day does not increase in line with the number of nurses. This can be explained by the fact that a triage team at night is staffed by one experienced nurse who can perform all activities in the triage process and on average cope with 3.5 patients per hour. During day time, a triage team is staffed by one experienced nurse and one less experienced nurse, who can perform a part of all triage activities. The triage team at day time serves on average 5.5 patients per hour. One conclusion to be drawn from this is that doubling the number of nurses does not necessarily lead to a doubled effective capacity in the triage process, as assumed in the

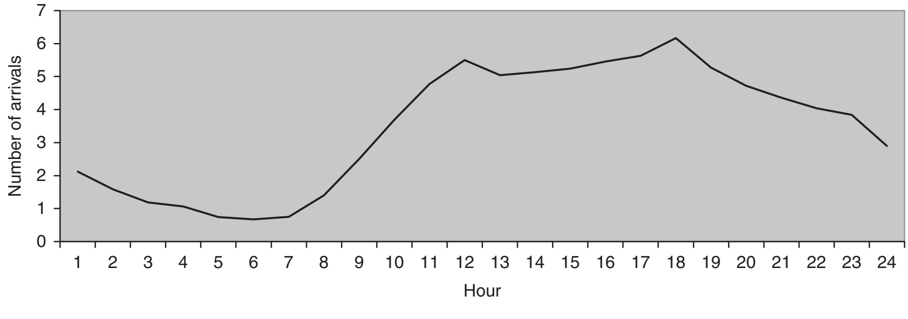

Figure 2. Average number of arrivals per hour to the emergency department during 2011

Hour	Queue model	λ Sample mean	W_q Sample mean	μ Estimated mean (per server)	Estimated mean (the system)
1	MM1	2.118	0.070	6.661	6.661
2	MM1	1.586	0.050	6.481	6.481
3	MM1	1.186	0.030	6.908	6.908
4	MM1	1.063	0.032	6.322	6.322
5	MM1	0.745	0.027	5.637	5.637
6	MM1	0.674	0.032	4.939	4.939
7	MM1	0.751	0.044	4.506	4.506
8	MM2	1.403	0.062	2.077	4.154
9	MM2	2.501	0.078	2.908	5.816
10	MM2	3.688	0.096	3.628	7.256
11	MM2	4.773	0.114	4.196	8.392
12	MM2	5.504	0.161	4.302	8.604
13	MM2	5.038	0.149	4.093	8.186
14	MM2	5.134	0.121	4.366	8.732
15	MM2	5.241	0.122	4.426	8.852
16	MM2	5.455	0.132	4.476	8.952
17	MM2	5.633	0.149	4.458	8.916
18	MM2	6.170	0.150	4.775	9.550
19	MM2	5.271	0.152	4.216	8.432
20	MM2	4.721	0.129	4.035	8.070
21	MM2	4.364	0.132	3.782	7.564
22	MM2	4.041	0.139	3.523	7.046
23	MM1	3.844	0.132	7.651	7.651
24	MM1	2.901	0.091	7.279	7.279

Table I. Estimated effective capacity measures

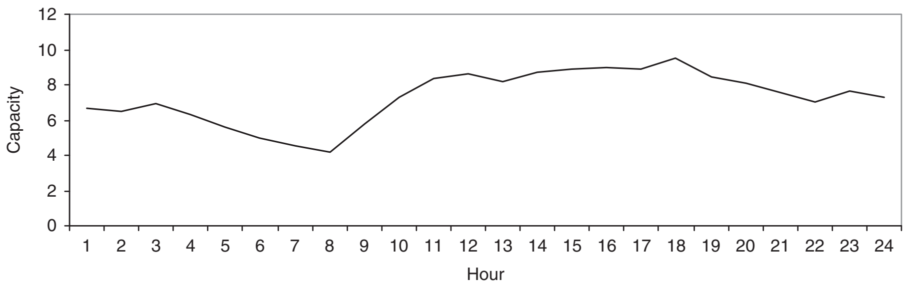

Figure 3. Estimated effective capacity

planning process at the emergency department. It is also interesting to note the slowly declining capacity at night. One explanation for this is that the nurse is not entirely dedicated to the triage, when the number of arrivals to the emergency department gradually decreases.

The conformity between effective capacity and number of nurses in Figure 4 seems to be worst during the transition hours between day time and night time, when two nurses are scheduled to work (i.e. between 7 a.m. and 11 a.m. and between 8 p.m. and 11 p.m.). From discussions with management, we have found that probable reasons for this is that day time nurses stay longer if demand is high or that nurses are borrowed for the triage process from other parts of the hospital to work away the queue. Hence, in practice, the actual average number of nurses increases (decreases) continuously from two to four (four to two) during the transition hours. The effect is clearly illustrated by significant bumps in the curve in Figure 5, in which hourly estimated average effective capacity per nurse is displayed, during the transition hours. A simple interpolation would probably provide better estimates of the effective capacity numbers during the transition hours.

In Figure 6, hourly estimated effective capacity in the triage process is compared to the hourly average number of arrivals. Here the coherence is strong, which indicates that the management of capacity is well adapted to the actual demand. Most of the time, the average utilization revolves around 0.5, which means that there is generally enough

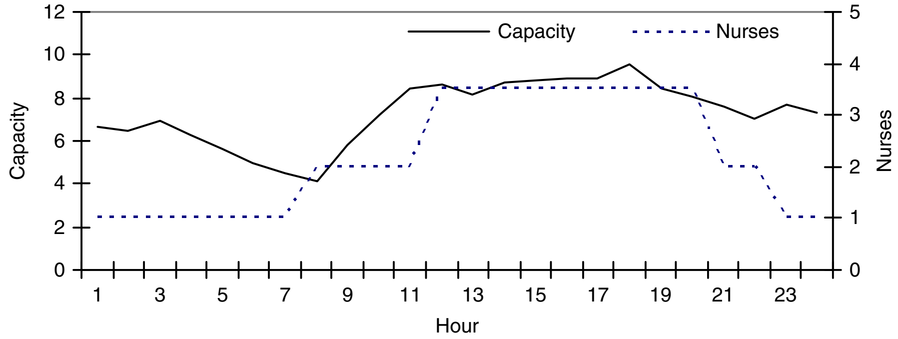

Figure 4.
Estimated effective capacity and actual number of nurses

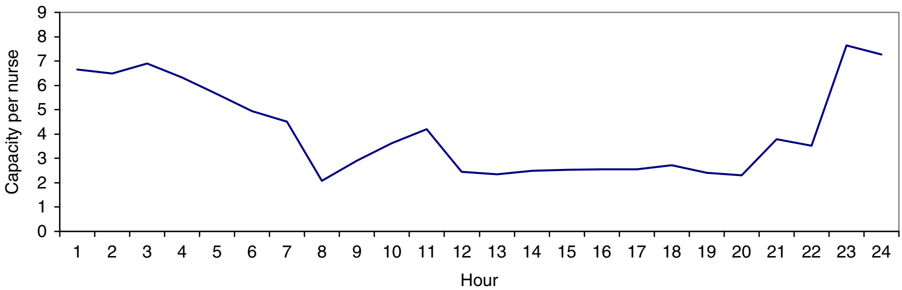

Figure 5.
Estimated average effective capacity per nurse

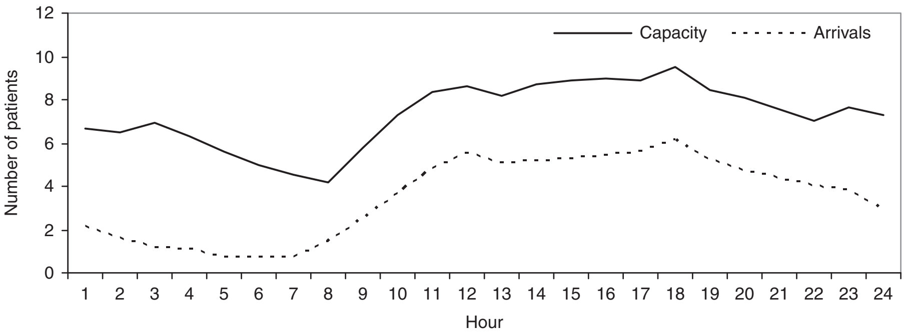

Figure 6.
Estimated effective capacity and average number of arrivals

capacity to deal with a demand up to twice the average level. However, during the late part of the night, the utilization is significantly lower since there cannot be less than one nurse working with triage, even during times when demand is low.

One problem that was addressed in discussions with the management of the emergency department is that because arrivals have a Poisson distribution, the variance in the number of arrivals varies during the day in the same way as the mean. Hence it is unreasonable to have a constant excess capacity in the triage if the required service levels are formulated in terms of a certain percentage of all patients meeting a nurse in the triage within a certain time, for example "90% of all patients shall meet the triage team within 10 minutes."

A constant excess capacity in the triage process would cause service levels to vary substantially during the day and night, and would be lowest when the number of arrivals is highest, as shown in Figure 7.

If the goal is to achieve an equally high service level during the day and night, the excess capacity must be dimensioned, given that the average demand (the number of arrivals) varies. Excess capacity will then be highest when the number of arrivals is highest.

4. Conclusions

Skaraborg Hospital is working on an increasingly advanced production and capacity planning system and is dependent on a reliable methodology for measuring effective capacity. According to the management of the emergency department, to date, the effective capacity of the emergency department has been estimated using conventional methods or by direct empirical observations. Hence, the proposed methodology in this paper could be an appropriate alternative to these traditional methods for production and capacity planning. The method can also be useful to evaluate the impact of implemented changes in work procedures, the staffing plan, and so on, in terms of effective capacity.

A number of highly relevant managerial conclusions can be drawn from the results of the capacity estimates. The first is that a doubling of the number of nurses in the triage does not double the effective capacity, contrary to common belief. The second is that the notion, among politicians and other decisions makers, that a general increase in resources is the standard solutions to many problems (Walley, 2013), such as long waiting times, in the emergency department is not true. Instead, what is important is where the additional resources will be delivered, when they will be delivered, and how much should be delivered to ensure a certain service level. The third conclusion is that

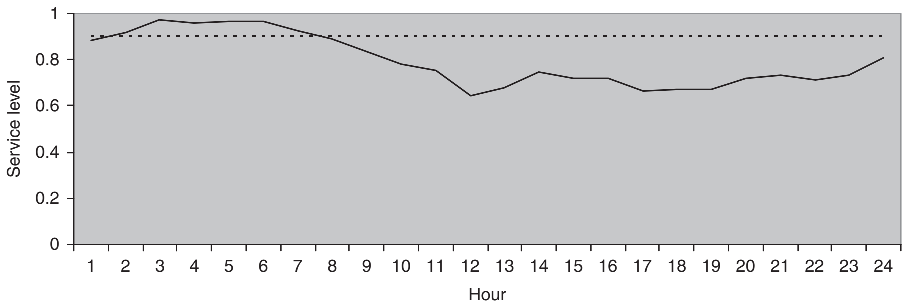

Figure 7. Actual average service levels if staffed with constant excess capacity

the excess capacity must be related to the arrival rates to achieve an equally high service level during a 24-hour period.

This paper contributes to the literature in this field in two different ways. First, it represents a contribution to the theory of capacity measurement by suggesting a methodology through which the effective capacity of a service process can be estimated objectively without direct observation, thereby also avoiding the potential risk of the Hawthorne effect. The method can be used to estimate the effective capacity of most service processes as long as the arrival process and the waiting times can be observed. Second, we conducted a case study with real data to illustrate how the methodology can be applied to obtain capacity measures for the operations management of a triage process in an emergency room.

One limitation when applying the method is that it only gives valid results for completely dedicated and standardized processes. In the case study, we can conclude that the calculation of effective capacity becomes misleading during those times of day when the nurses in the triage process perform duties that are not part of the triage process. The corresponding measurement error occurs even when the nurses in the triage process does not perform all stages in the triage process.

Another limitation of our results is that the assumption of exponential service times is very unlikely to be true in practice. However, it is a common assumption in similar studies where queueing models are used to analyze capacity in health care settings (e.g. Green, 2006, 2011).

Future research in this area should be directed toward implementation of this methodology in other types of processes. As long as the arrival process and the waiting times can be observed, and the assumptions of a relevant queueing model are valid, the idea can be used to estimate the effective capacity of any service process or bottleneck. We see many possible applications, particularly in operations and logistics management. However, to analyze more complex flows, more advanced queueing models need to be used. Hence, for capacity measurement, the utilization of queueing models other than the ones covered here should also be investigated.

References

Balakrishnan, N., Render, B. and Stair, R. (2007), *Managerial Decision Modeling with Spreadsheets*, Prentice Hall, Upper Saddle River, NJ.

Bamford, D. and Chatziaslan, E. (2009), "Healthcare capacity measurement", *International Journal of Productivity and Performance Management*, Vol. 58 No. 8, pp. 748-766.

Bowers, J.A. (2011), "Simulating waiting list management", *Health Care Management Science*, Vol. 14 No. 3, pp. 292-298.

Bratt, J.H., Foreit, J., Chen, P.-L., West, C., Janowitz, B. and De Vargas, T. (1999), "A comparison of four approaches for measuring clinician time use", *Health Policy Plan*, Vol. 4 No. 4, pp. 374-381.

Campbell, J.P., Maxey, V.A. and Watson, W.A. (1995), "Hawthorne effect: implications for prehospital research", *Annals of Emergency Medicine*, Vol. 26 No. 5, pp. 590-594.

De Bruin, A.M., van Rossum, A.C., Visser, M.C. and Koole, G.M. (2007), "Modeling the emergency cardiac in-patient flow: an application of queueing theory", *Health Care Management Science*, Vol. 10 No. 2, pp. 125-137.

Finkler, S.A., Knickman, J.R., Hendrickson, G., Lipkin, M. and Thompson, W.G. (1993), "A comparison of work-sampling and time-and-motion techniques for studies in health services research", *Health Services Research*, Vol. 28 No. 5, pp. 577-597.

Fitzsimmons, J.A. and Fitzsimmons, M.J. (2008), *Service Management – Operations, Strategy, Information Technology*, McGraw-Hill, New York, NY.

Franke, R.H. and Kaul, J.D. (1978), "The Hawthorne experiments: first statistical interpretation", *American Sociological Review*, Vol. 43 No. 5, pp. 623-643.

Green, L.V. (2006), "Queueing analysis in healthcare", in Hall, R.W. (Ed.), *Patient Flow: Reducing Delay in Health Care Delivery*, Springer, New York, NY, pp. 281-307.

Green, L.V. (2011), "Queueing theory and modeling", in Yih, Y. (Ed.), *Handbook of Healthcare Delivery Systems*, Taylor & Francis, London, pp. 1-16.

Green, L.V., Kolesar, P.J. and Soares, J. (2001), "Improving the SIPP approach for staffing service systems that have cyclic demand", *Operation Research*, Vol. 49 No. 4, pp. 549-564.

Green, L.V., Soares, J., Giglio, M.D. and Green, M.D. (2006), "Using queueing theory to increase the effectiveness of emergency department provider staffing", *Academic Emergency Medicine*, Vol. 13 No. 1, pp. 61-68.

Gross, D., Shortle, J.F., Thompson, J.M. and Harris, C.M. (2008), *Fundamentals of Queueing Theory*, John Wiley & Sons, Hoboken, NJ.

Izady, N. and Worthington, D. (2012), "Setting staffing requirements for time dependent queueing networks: the case of accident and emergency departments", *European Journal of Operational Research*, Vol. 219 No. 3, pp. 531-540.

Koeleman, P.M., Bhulai, S. and van Meersbergen, M. (2012), "Optimal patient and personnel scheduling policies for care-at-home service facilities", *European Journal of Operational Research*, Vol. 219 No. 3, pp. 557-563.

Kuntz, L., Scholtes, S. and Antonio, V. (2007), "Incorporating efficiency in hospital-capacity planning in Germany", *European Journal of Health Economics*, Vol. 8 No. 3, pp. 213-223.

Laffel, G. and Blumenthal, D. (1989), "The case for using industrial quality management science in health care organizations", *Journal of the American Medical Association*, Vol. 262 No. 20, pp. 2869-2873.

Lakshmi, C. and Sivakumar, A.I. (2013), "Application of queueing theory in health care: a literature review", *Operations Research for Health Care*, Vol. 2 Nos 1-2, pp. 25-39.

Li, D. and Glazebrook, K.D. (2011), "A Bayesian approach to the triage problem with imperfect classification", *European Journal of Operational Research*, Vol. 215 No. 1, pp. 169-180.

Mayhew, L. and Smith, D. (2008), "Using queueing theory to analyse the government's 4-h completion time target in accident and emergency departments", *Health Care Management Science*, Vol. 11 No. 1, pp. 11-21.

Mehdi, J. (2002), *Stochastic Models in Queueing Theory*, Academic Press, New York, NY.

Roethlisberger, F.J. and Dickson, W. (1939), *Management of the Worker*, Harvard University Press, Cambridge, MA.

Russell, R.S. and Taylor, B.W. (2006), *Operations Management – Quality and Competitiveness in a Global Environment*, John Wiley & Sons, Hoboken, NJ.

Sasser, W.E. (1976), "Match supply and demand in service industries", *Harvard Business Review*, Vol. 54 No. 6, pp. 133-140.

Schroeder, R.G. (1993), *Operations Management – Decision Making in the Operations Function*, McGraw-Hill, New York, NY.

Sittig, D.F. (1993), "Work-sampling: a statistical approach to evaluation of the effect of computers on work patterns in healthcare", *Methods of Information in Medicine*, Vol. 32 No. 2, pp. 167-174.

Slack, N., Chambers, S. and Johnston, R. (2010), *Operations Management*, Financial Times/ Prentice Hall, Harlow.

Walley, P. (2013), "Does the public sector need a more demand-driven approach to capacity management?", *Production Planning & Control*, Vol. 24 Nos 10-11, pp. 877-890.

Wickström, G. and Bendix, T. (2000), "The 'Hawthorne effect' – what did the original Hawthorne studies actually shows?", *Scandinavian Journal of Work, Environment and Health*, Vol. 26 No. 4, pp. 363-367.

Yen, K., Shane, E.L., Pawar, S.S., Schwendel, N.D., Zimmanck, R.J. and Gorelick, M.H. (2009), "Time motion study in a pediatric emergency department before and after computer physician order entry", *Annals of Emergency Medicine*, Vol. 53 No. 4, pp. 462-469.

About the authors

Björn Lantz is an Associate Professor of Engineering Economic Analysis at the Chalmers University of Technology. He has a PhD in Business Administration from the University of Gothenburg (2000). His research interests include health care, operations management, logistics, and methodology.

Peter Rosén is an Associate Professor of Business Administration at the University of Gothenburg. He has a PhD in Business Administration from the University of Gothenburg (1992). His research interests include health care, operations management, and logistics.