Determining the optimal capacity and occupancy rate in a hospital: a theoretical model using queuing theory and marginal cost analysis

Authors:

1. Muhammad Maaz*[1,2]

   maazm@mcmaster.ca

   647-712-4251

   ORCID: 0000-0002-3869-631X

2. Anastasios Papanastasiou[2]

   papanasa@mcmaster.ca

   905-525-9140, ext. 24593

Affiliations:

1. Faculty of Health Sciences, McMaster University

   1280 Main St. W., Hamilton, ON, Canada L8S4L8

2. Department of Economics, McMaster University

   1280 Main St. W., Hamilton, ON, Canada L8S4L8

*corresponding

This is the pre-peer reviewed version of an article published in Managerial and Decision Economics, which has been published in final form at https://doi.org/10.1002/mde.3176.

Abstract: We model the hospital as seeking to balance the costs to itself in providing care, as well as the societal cost of people waiting for care. We use queuing theory to show that the optimal capacity and the corresponding optimal occupancy rate are dependent on the marginal cost of expanding capacity, the marginal cost of waiting, and the rates of patient arrival and discharge. Therefore, a universal occupancy target is unfounded. As well, the model shows that increasing capacity to respond to increased patient influxes is inadequate, suggesting that the healthcare system must explore alternate responses to burgeoning patient populations.

1. Introduction

Occupancy rates in hospitals are a major concern of health policy. There is ample evidence that high occupancy rates are intimately associated with poor outcomes such as greater risk of premature death and readmission (Chrusch et al., 2009). In a systematic review, Kaier et. al. (2012) showed that high bed occupancy rates directly influenced the incidence of hospital-acquired infections. Occupancy rates have been a great matter of concern for many healthcare systems, especially those with universal healthcare, for example Canada. While exact statistics on the occupancy levels are largely unavailable, there are some indications as to the extent of the problem. According to an audit by the Ontario Ministry of Health, 60% of all medicine wards in Ontario's large community hospitals have occupancy rates higher than 85%. Indeed, overcrowding in Ontarian hospitals has somewhat become the new "norm", as figures show multiple hospitals with higher than 100% occupancy (Grant 2017). The target of 85% is notable, as this is the number that the Ministry looks to in order to guide its hospital capacity determinations. The 85% target seems to be a widely accepted target, and is referenced widely in various other governmental publications, including those by the UK Department of Health (2017). The general justification for this occupancy rate is to limit unused resources, while still maintaining some excess capacity to accommodate surges of patient influx. As governments continue to set standards for hospital capacity in the future, a rigorous determination of the optimal capacity, which is a crucial determinant of the occupancy rate, is key.

The question of an optimal capacity for hospitals has a rich literature associated with it. Numerous previous papers have used results from queuing theory, the mathematical study of wait lines or queues (Gnedenko 1989), in order to calculate the optimal number of beds given certain criteria. For example, Jones (2011) looked at the required number of beds for various levels of allowable wait times and various allowable probabilities of an incoming patient being forced to wait. Similarly, Gorunescu et. al. (2002) used models from queuing theory in order to simulate how a hospital would function. However, they used a loss model, meaning they assumed that any incoming patient who did not receive immediate service would simply give up and go home. However, this model is not fully realistic, as surely prior knowledge tells us that patients do indeed wait for service – for beds, for procedures, etc. These and other studies show that hospital operations are ripe for analysis through the lens of queuing theory.

However, we argue that prior applications of queuing theory to hospital operations have been lacking in two major ways. Firstly, there is often no consideration of the cost of expansion. Surely, to increase capacity costs resources: not just the money to build the new bed, but also staffers and the space it occupies. Papers that determine the optimal number of beds based solely on functional or health related outcomes like average wait time implicitly assume that coffers are limitless and that we can build however many beds we need in order to fulfill such criteria. This is clearly not a realistic assumption, and a consideration of costs is necessary to make such models more realistic, seeing as spending on hospital beds necessarily elicits a trade-off in spending elsewhere. Secondly, the mathematical treatment of such models has been cursory. For example, Kembe et. al. (2012) did incorporate cost considerations into their queuing theory results, but only applied the equations to one specific hospital. This is really a rather routine numerical application, and unfortunately, generalization to various levels of hospital characteristics was lacking. Dynamic shifts in variables were not considered, and scenarios like surges of patients and how this would consequently affect the functioning of the hospital were not explored.

Our paper is novel because, while previous papers have explored the issue of the optimal number of beds from a purely health outcome perspective, we also incorporate the costs of expansion into the determination of optimal capacity. We then consider changes in the constituent variables of the model and see what the model can tell us about real-life policy regarding guidelines for hospital capacity. Lastly, we discuss how governments ought to respond going forward to likely future changes in the healthcare system.

2. Methods

2.1. A sketch of the development of the model

The model is roughly developed as follows. We assume that hospitals face two costs: the cost it incurs from maintaining a certain number of beds, and the cost of patients waiting. The hospital wishes to balance these two costs, so as to minimize the total cost. We define a total cost function incorporating both of these considerations. Next, we minimize this cost function in order to arrive at the optimal capacity. We then extend this optimal capacity level to find the corresponding occupancy rate, through another application of queuing theory. Lastly, we perform analyses on how various factors and situations affect the optimal capacity and hence the occupancy rate.

2.2. The total cost function

At every level of capacity c, there is a cost associated with purchasing and maintaining that number of beds. We call this the cost of service $S(c)$, and we assume that it is a linear function such that $S(c) = Ac$, where A is simply the per-bed cost. We would also like to quantify the benefit gained by having beds; however, we will do this by instead turning the problem on its head and looking at the cost of not having beds. This is similar to an approach used previously by Kembe et. al. (2012). We therefore define $W(c)$, the cost of waiting, and define it to be proportional to the number of people waiting. In a way, this cost function is a quantification of the adverse health outcomes associated with waiting. We therefore say that $W(c) = BL(c)$, where B is the average cost of waiting, and $L(c)$ is the number of people in the queue, as a function of capacity. We note here that $W'(c) < 0$, which is intuitive due to the expected decrease in the size of the queue as we increase the number of beds. With these two functions, we can construct our total cost function $C(c)$, which is simply the sum of the cost of service and the cost of waiting.

$$C(c) = S(c) + W(c)$$

2.3. Optimization condition of $C(c)$

Differentiating with respect to the level of capacity,

$$C'(c) = S'(c) + W'(c)$$

Hence, for every marginal increase in the number of beds c in a hospital, there are two factors at play: the increased cost due to adding that bed, and the decreased cost of waiting. The first order optimization condition of the function is simply setting $C'(c) = 0$, and hence we have the important statement that to optimize capacity, we must choose it such that

$$S'(c) = -W'(c)$$

Therefore, the optimal capacity is when the marginal cost of adding that bed is balanced by the marginal decreased cost of waiting. As noted above, $W'(c) < 0$, while $S'(c) > 0$ and is constant. We are only interested in the interior solutions to this optimization condition, although we note the possible existence of corner solutions as well[1].

2.4. Defining the marginal service cost function $S'(c)$

Since $S(c) = Ac$, we have that

$$S'(c) = A$$

2.5. Defining the marginal waiting cost function $W'(c)$

Based on the definition of $W(c)$, it is first necessary to understand $L(c)$, the number of people in the queue, before we can deduce $W'(c)$. For this, we turn to queuing theory, a field of mathematics which models the formation of queues. Based on certain pieces of information about the system – in our case, the hospital ward – we can describe various properties of the queue.

[1] If $|S'(c)| > |W'(c)|$ for all c, then decreasing c will always decrease $C(c)$, and the cost minimizing capacity level is to have no beds. On the other hand, if $|W'(c)| > |S'(c)|$ for all c, then the optimal capacity is to increase the number of beds without bound, or until we reach some stipulated upper bound, as for every bed added, the cost of adding the bed is smaller than the cost saved (i.e. benefit) of decreasing waiting and related adverse outcomes.

2.6. Using queuing theory to model the hospital

First, let us examine how the hospital ward operates, in an ideal situation. There are c number of beds to which incoming patients, arriving randomly and independently of each other at an average rate per unit time λ, may be allocated. This assumption may not be appropriate in some situations, for example an epidemic event, where certainly these arrivals are not independent of one another – but, we argue that in regular day-to-day functioning, we can assume independent arrivals. Patients spend a certain time occupying whatever bed they are in, and then are discharged, such that patients are discharged from the hospital at an average rate μ. We will assume that all the beds are identical, so that we do not care to which bed a certain patient is allocated. Furthermore, only one queue is formed: all the patients are in one line awaiting a bed; there are not individual queues for each bed. We will further assume that patients are allocated to a bed in a first come first serve basis: no prioritization of patients is performed. Based on these simple assumptions, we model both arrivals and discharges as Poisson variables, and define their average or expected value as λ and μ respectively. Empirical evidence justifies this claim, as analyses of hospital data have shown that both arrivals and discharges in various wards are well-approximated by a Poisson distribution, including obstetrics (Gam et. al., 2013; Gao et. al. 2017) and emergency (Whitt and Zhang, 2017).

This is important, as this means that the hospital can be appropriately modelled with a type of queuing model called an M/M/c queuing model. The M refers to the Poisson nature of λ and μ, while the c refers to the capacity of the hospital as we have discussed throughout this paper. Given the three variables λ, μ, and c, the M/M/c model provides us with the following important results. The average occupancy rate ρ is given as follows (Zukerman, 2013):

$$\rho = \frac{\lambda}{c\mu}$$

This formula aligns with our intuition that the occupancy rate increases along with the arrival rate and falls with increases in capacity or the rate of discharge. The expected number of people in the queue L is (Zukerman 2013):

$$L(c) = P_0\left(\frac{\rho^{c+1}}{(c-1)!\,(c-\rho)^2}\right)$$

P_0 is the probability of having zero patients in the ward, and is given by [13]:

$$P_0 = \left[\sum_{i=0}^{c-1}\frac{\rho^i}{i!} + \frac{\rho^c}{c!\,(1-\rho)}\right]^{-1}$$

We observe that as $\rho = {}^{\lambda}/_{c\mu}$, we can redefine $L(c)$ as a function of c and ${}^{\lambda}/_{\mu}$. For ease of notation, we declare the variable ε, and define it as:

$$\varepsilon = \frac{\lambda}{\mu}$$

The interpretation of ε is that it is the ratio of arrival rate to discharge rate, or the inverse turnover rate. Hence, we rewrite the queue length function as $L(c,\varepsilon)$. Having rigorously treated L, we return to our discussion of the marginal waiting cost function.

2.7. Returning to the marginal waiting cost $W'(c)$

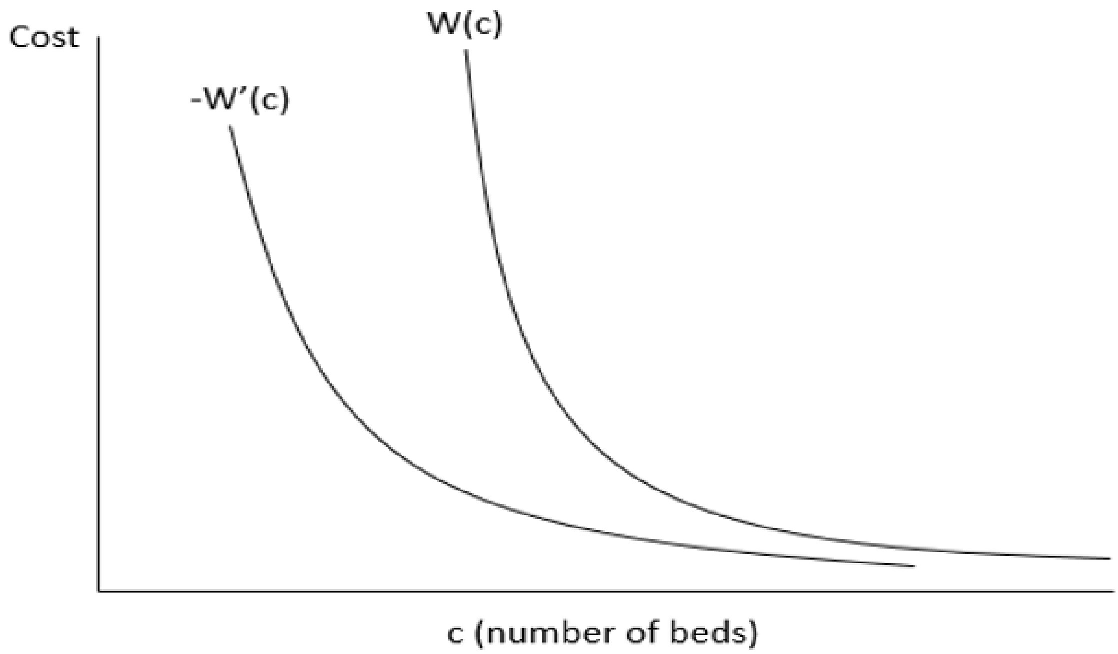
From our discussion above about $L(c, \varepsilon)$, we see that $W(c)$ is also dependent on ε, and so we write:

$$W'(c, \varepsilon) = BL'(c, \varepsilon)$$

However, fixing a certain value of ε yields a waiting cost function that is only dependent on c. This leads to an important insight: we must fix a certain ε if we seek to arrive at an optimal level of capacity c^*; or, said in another way, c^* is a function of ε. Our use of ε also leads to an intriguing conclusion: the actual values of λ or μ do not matter – it is rather their ratio that is important. In this way, a high λ does not necessarily imply a high c^*, as we need to look at it relative to μ.

Unfortunately, the formula for L is arduous to look at and so it is not immediately clear how $W(c)$ looks. More importantly, it is non-differentiable, as it is defined only for discrete inputs, due to the factorial term. Hence, we cannot use standard calculus techniques to describe $W(c)$ and $W'(c)$. We instead analyse it from a computational point of view. The pseudocode is attached in Appendix 1, but we describe our approach here in plain terms. We generated a $L(c)$ curve by holding ε constant and graphing $L(c)$ against c for a range of c values. Note that $W(c)$ is simply $L(c)$ scaled by some factor B, and so knowing the general behaviour of the $L(c)$ curve also tells us about the $W(c)$ curve. We also took very tiny marginal increases in c when generating $L(c)$, thereby mimicking infinitesimal changes of c, to generate the $W'(c)$ curve. Through these techniques, we arrived at the following graphical depiction of the curves.

Figure 1. Shapes of waiting cost function and marginal waiting cost function

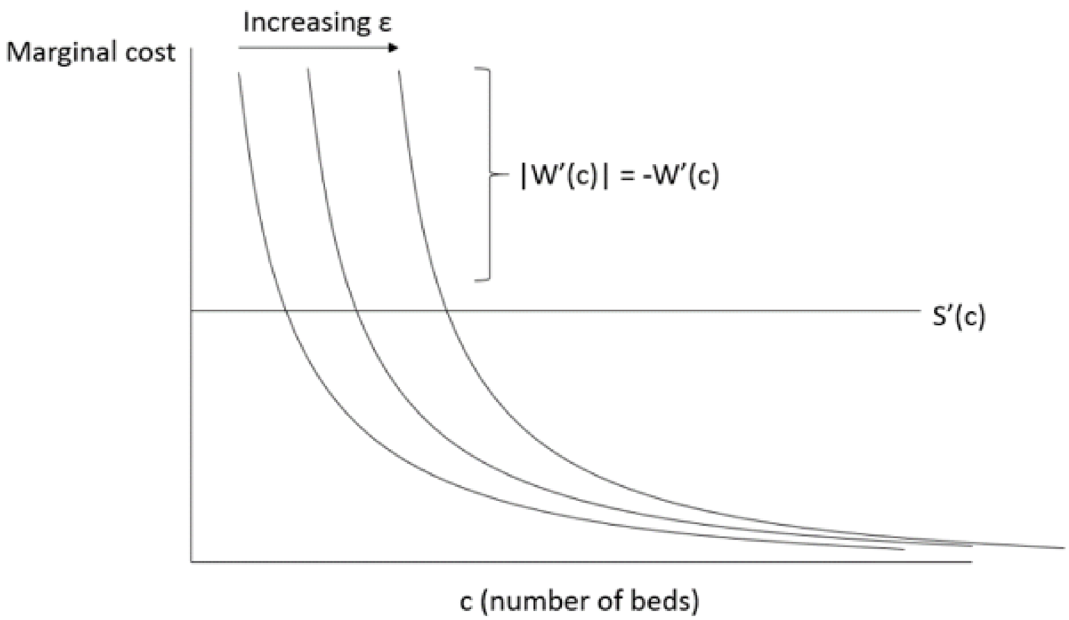


Interestingly, we found that $W(c)$ is well-approximated by an exponential decay function, and so it follows from calculus that $W'(c)$ should be as well. We observe that increasing ε shifts the $-W'(c)$ curve to the right, and, vice versa, a decrease shifts the curve to the left. Intuitively, this comes from the fact that as ε increases (an increase in arrivals or a decrease in discharges), we require a greater number of beds to maintain the same queue size. As the marginal cost of waiting is proportional to the length of the queue, we require a greater number of beds at any given level of marginal cost of waiting.

2.8. Optimizing the number of beds

We can depict the optimal number of beds as the intersection of the $S'(c)$ curve and the $-W'(c)$ curve. As $S'(c) = A$, it is simply a flat line, and we have already drawn $-W'(c)$ above in Figure 1.

Figure 2. Determining the optimal number of beds



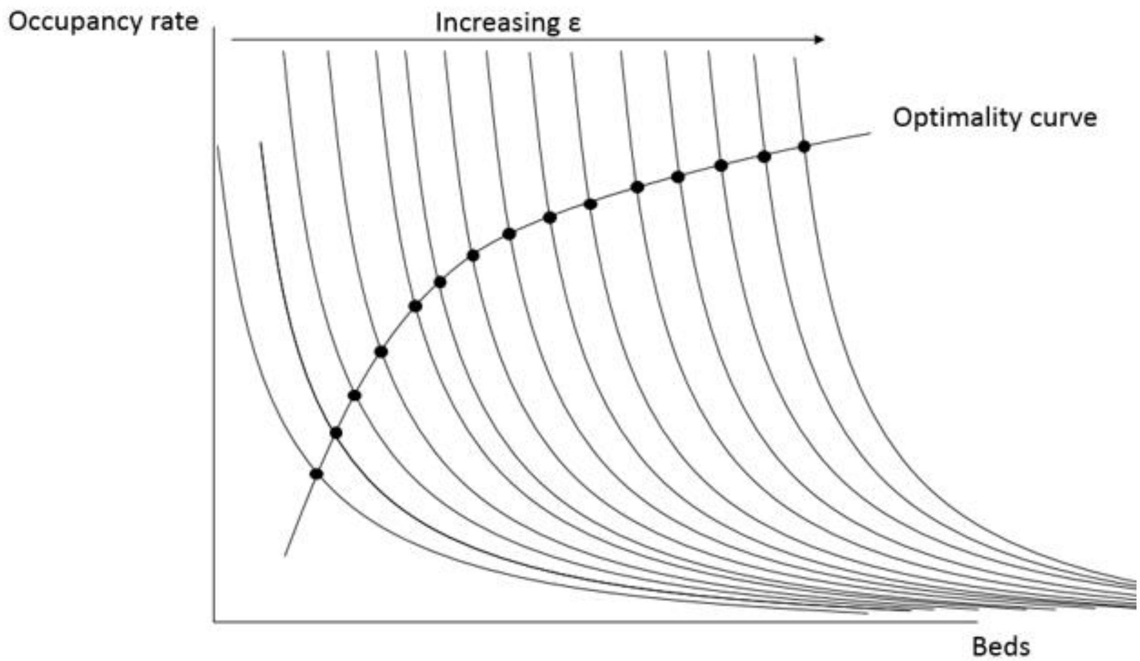
In Figure 2 above, we see that there is a unique optimal capacity, which can be found by looking at the intersection of the two curves for a given ε, as for each ε, there is a unique $-W'(c)$ curve. Furthermore, we see that increasing ε increases the optimal capacity. This is a rather intuitive result if we recall that $\varepsilon = {}^{\lambda}/_{\mu}$. For example, if more patients arrive, so that λ and hence ε increase, then it is quite sensible that we should need more beds in order to mitigate the higher costs of waiting. Similarly, if fewer patients are being discharged, so μ decreases and so ε increases, then it is also similarly intuitive that we would need more beds as there are more patients in the system. The model exactly matches our intuition as to how these factors would affect the optimal number of beds, which in turn justifies our assumptions and modelling thus far.

2.9. Extending the model to occupancy rate ρ

We previously gave a definition of the average occupancy rate as $\rho = {}^{\lambda}/_{c\mu}$, which we can also write as $\rho = {}^{\varepsilon}/_{c}$, since $\varepsilon = {}^{\lambda}/_{\mu}$. The relationship between c and ρ is injective, so that, for a given ε, there exists a unique optimal occupancy rate for a given number of beds. We are interested in how changing the properties of the hospital in terms of arrivals and discharges affects the occupancy rate. Looking at Figure 2, if we increase ε, it is optimal to increase c. However, as $\rho = {}^{\varepsilon}/_{c}$, it is ambiguous, looking from just a graphical point of view, as to how a change in ε affects ρ, as this depends on the relative magnitudes of the changes in ε and c. Again, we are limited in our understanding by the non-differentiability of $W(c)$, and once again turn to computational techniques to better understand how occupancy rates are affected. Our full pseudocode is provided in Appendix 1.

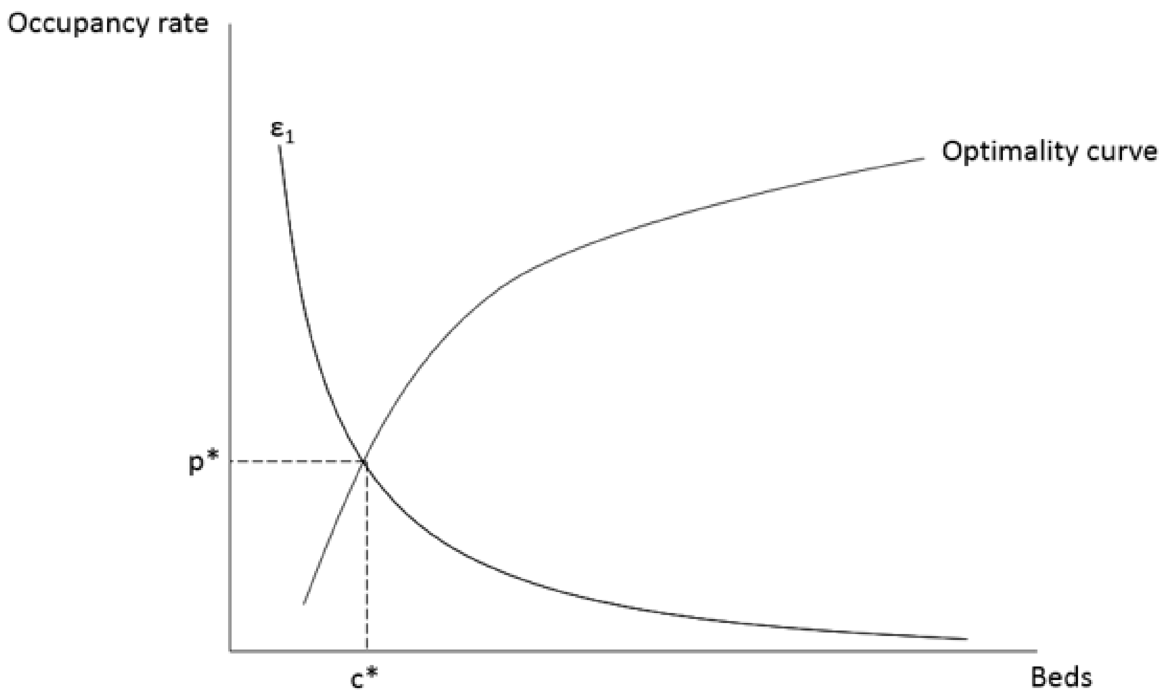
We sought to depict this relationship by graphing ρ vs c. Given that $\rho = {}^{\varepsilon}/_{c}$, if we fix ε, we get a unique curve showing the occupancy level at each level of capacity. Exactly one point on that curve corresponds to the optimal number of beds and the corresponding optimal capacity (c^*, ρ^*). Essentially what we sought to do was "connect the dots": draw a curve connecting the optimal points for all ε. We allowed ε to vary over a large range of values, stepping by very small increments. We generated a $W'(c)$ curve for each level of ε, using the method described previously. Note that we simply arbitrarily chose $A = B = 1$ initially, as we are interested in the general behaviour, and A and B are simply scaling factors. Indeed, we later found that our practical results are robust to this assumption. We then found the intersection with the $S'(c)$ to arrive at the optimal capacity c^*, and the corresponding occupancy rate ρ^*. We then connected all the pairs (c^*, ρ^*) generated from varying ε, and thus constructed an optimality curve. The visualization of this approach is shown in Figure 3.

Figure 3. Generating the optimality curve by connecting the optimality points of occupancy curves for a wide range of ε



Our optimality curve is increasing, concave, and approaches $\rho = 1$. Its property as unambiguously increasing is surprising. We also observe that such an optimality curve is not stationary and can change depending on A and B. We note that any point on this optimality curve is also a point on the curve of ρ vs c for a given level of ε. Therefore, given an optimality curve, if we graph ρ vs c for a fixed ε, the intersection of the two curves yields the optimal capacity and optimal occupancy for that level of ε. This is shown below in Figure 4.

Figure 4. The relationship between occupancy rate and level of capacity

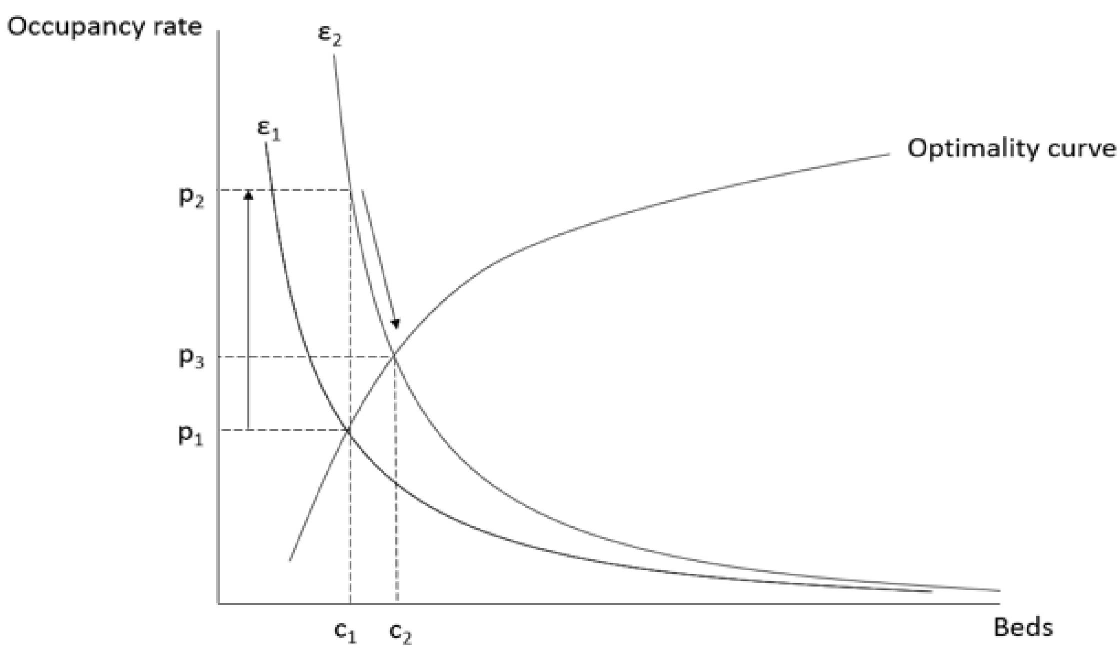


3. Results

We now explore the consequences of Figure 4. First, we look at the case with a constant ε. If a hospital is operating at too high of a capacity, beyond its c^*, then it will experience lower occupancy rates. This is not necessarily good, as the hospital is not operating on the optimality curve: at this point, the marginal cost of service outweighs the marginal decreased cost of waiting. It may optimize itself by decreasing the number of beds, although this will mean an increase in occupancy rates. Conversely, if the hospital has too few beds, such that it has less than c^* beds, then it will experience too-high occupancy rates. In this case, the hospital would benefit from increasing its number of beds, as this would allow it to cost-minimize and decrease occupancy rates.

Now we consider what would happen if we allow ε to change. We already know from Figure 2 and from our intuition that as ε increases, c should increase to match it. Rather surprisingly, according to Figures 3 and 4, it also increases the occupancy rate ρ (and likewise, for a decrease in ε, ρ decreases). We can model this by saying we move from a value of ε_1 to ε_2 – that is, the rate of arrivals relative to the rate of discharge increases. In response to this increase in incoming patients, the cost-minimizing hospital responds by increasing the number of beds from c_1 to c_2, so that it can continue to operate optimally. Alas, the occupancy rate also, rather counterintuitively, increases as well. This is depicted below.

Figure 5. The hospital's response to an epsilon shock



The increase in ε is modelled as the shift from the ρ-curve of ε_1 to that of ε_2. In the short run, the hospital is still operating at the same number of beds c_1 and so its occupancy rate increases from ρ_1 to ρ_2. In the long run, the hospital attempts to move back to the optimality curve, and so increases its capacity from c_1 to c_2, and therefore decreases its occupancy rate to ρ_3. With an increase in arrivals relative to discharges, the occupancy rate increases, as is intuitive. In response to this the hospital increases its capacity, and though this decreases the occupancy rate from that initial shock, it is yet still larger than the occupancy rate that was present before the increase in arrivals. As we will further explore in the discussion, this suggests that increasing the number of beds to accommodate increased patient influx is not necessarily a wholly beneficial policy – though we are still cost-minimizing, we yet experience an increased occupancy rate.

4. Discussion

Numerous insights into hospital capacity and occupancy rates may be gleaned from Figure 4 and Figure 5. Firstly, a one-size-fits-all conception of occupancy rate is unfounded. As we discussed in the introduction, a target of 85% is widely cited in government policy. However, what Figure 4 shows is that there is in fact a range of optimal occupancy rates, and the optimal one for a specific hospital depends on the hospital's individual characteristics (namely, its ε, and even its A and B). Attempting to push a single occupancy target onto every hospital without consideration of the variance of these characteristics is not ideal.

The second important result we see from our discussion of changes in ε is that despite increasing the number of beds in the long term to remain on the optimality curve as a response to an increase in ε, the occupancy rate yet increases. Despite the hospital's efforts, the percentage of its beds that are occupied increases. This is a result that runs directly counter to a prevailing notion of health policy and indeed counter to our instinctual understanding. Increased number of patients is often met with a recommendation to increase the capacity of the hospital, so as to mitigate the increase in occupied beds. However, what this model suggests is that this is a futile effort – the occupancy rate will continue to increase. The hospital is essentially fighting a losing battle, since indeed it cannot decrease the occupancy rate by responding in this fashion, as shown in Figure 4. Simply increasing capacity is an insufficient response if we believe that low occupancy rates are desirable. This result is of paramount importance as factors like an aging population will contribute to increased patient arrivals in the future (Canadian Medical Association 2013). We can conclude, therefore, that while the optimizing hospital does indeed mitigate the increase in occupancy resulting from an increase in arrivals, it alas cannot completely reverse it, and is doomed to experience, even in the long run after adjusting itself towards optimality, a higher occupancy rate than before.

While the model gives us policy criticisms, it also yields useful policy prescriptions. Continuing our exploration of the situation discussed above, we must stage another response to an increasing λ. A range of policy responses are possible simply by examining each of the constituent variables in the model. For example, previously we took ε to be in some way exogenous. However, it is quite possible that the healthcare system may be able to affect ε by modifying μ. As μ is the rate of discharge, this is in some way related to the efficiency of the hospital. Examination of current bureaucratic structures in healthcare, and aiming to cut down on these, would increase μ so as to balance the increase in λ (indeed, to match an increase in λ with the proportionally same increase in μ would keep ε constant). It is even arguable that health policy could affect λ: an increased emphasis on preventative care could mitigate the number of patients that arrive at the hospital. In the realm of the emergency ward, where concerns about high occupancy rates are perhaps most prevalent, pushing the notion that only actually emergent cases should present to emergency would be a positive step towards limiting overcrowding.

The optimality curve is also not necessarily static. By decreasing A, the per-bed cost, we shift the $S'(c)$ curve down, and hence the optimality curve would shift to the right to show that for a given ε, the optimal capacity is higher and in fact we yield a correspondingly lower occupancy rate. Therefore, in the situation depicted in Figure 4, if we responded by shifting the optimality curve by the necessary amount, then we could still be cost-optimizing and maintain, or even decrease, the occupancy rate. Extensive analysis is possible on each and every variable used in this model, each of which yield a multitude of associated policy measures that we can use to respond to increased number of patients, other than merely increasing capacity.

5. Conclusion

In this paper, we developed a cost-minimizing model that shows how one can determine the optimal level of capacity for a hospital. We later extended our analysis into occupancy rates, and, through computational techniques, derived surprising relationships between occupancy rates and hospital characteristics. Notably, there is no single optimal occupancy rate, but rather it is dependent on the hospital. We also showed that merely increasing the number of beds to match an increase in patients will not maintain low occupancy rates. Despite operating at an optimizing capacity, hospitals will be subject to increases in occupancy in the future. If keeping occupancy rates low is desirable, then the healthcare system must look to other ways of dealing with burgeoning patient populations.

Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

Canadian Medical Association. Health and Health Care for an Aging Population. December 2013.

Chrusch, Carla A., Kendiss P. Olafson, Patricia M. McMillan, Daniel E. Roberts, and Perry R. Gray. "High occupancy increases the risk of early death or readmission after transfer from intensive care." Critical care medicine 37, no. 10 (2009): 2753-2758.

Gam, Christiane MB, Julien Tanniou, Niels Keiding, and Ellen L. Løkkegaard. "A model for the distribution of daily number of births in obstetric clinics based on a descriptive retrospective study." BMJ open 3, no. 8 (2013): e002920.

Gao, Cheng, Abel N. Kho, Catherine Ivory, Sarah Osmundson, Bradley A. Malin, and You Chen. "Predicting length of stay for obstetric patients via electronic medical records." Studies in health technology and informatics 245 (2017): 1019.

Gnedenko, By, and I. N. Kovalenko. Introduction to Queuing Theory. Mathematical Modeling. Birkhaeuser Boston, Boston, 1989.

Gorunescu, F., S. I. McClean, and P. H. Millard. "A Queueing Model for Bed-Occupancy Management and Planning of Hospitals." The Journal of the Operational Research Society 53, no. 1 (2002): 19-24. http://www.jstor.org/stable/822875.

Grant, Kelly. "Hospital Overcrowding Has Become the Norm in Ontario, Figures Show." The Globe and Mail, May 21, 2017.

Jones, Rod. "Hospital bed occupancy demystified." British Journal of Healthcare Management 17, no. 6 (2011): 242-248.

Kaier, K., N. T. Mutters, and U. Frank. "Bed occupancy rates and hospital-acquired infections—should beds be kept empty?." Clinical Microbiology and Infection 18, no. 10 (2012): 941-945.

Kembe, M. M., E. S. Onah, and S. Iorkegh. "A study of waiting and service costs of a multi-server queuing model in a specialist hospital." International Journal of Scientific & Technology Research 1, no. 8 (2012): 19-23.

Ontario. Ministry of Health. Large Community Hospital Operations. 429-79. http://www.auditor.on.ca/en/content/annualreports/arreports/en16/v1_308en16.pdf.

United Kingdom. Department of Health. National Institute for Health and Care Excellence. Chapter 39 Bed Occupancy. July 2017.

Whitt, Ward, and Xiaopei Zhang. "A data-driven model of an emergency department." Operations Research for Health Care 12 (2017): 1-15.

Zukerman, M. (2013). Introduction to queueing theory and stochastic teletraffic models. arXiv preprint arXiv:1307.2968.

Appendix

Appendix 1. Pseudocode used to generate curves

We used R, which implements the M/M/c queuing model in its queuing package.

Loop ε from 0.1 to 50 stepping by 0.1

Loop c from minimum allowable beds[2] to 100,000 stepping by 1

Calculate L(c) for each c using the queuing package

Calculate W(c) by W(c) = B*L(c)

Calculate W'(c) by ΔW(c)

Store[3] (c, W(c)) in a list of ordered pairs

Store[4] (c, W'(c)) in a list of ordered pairs

Fit an exponential approximation[5] to the W(c) curve in the form m*exp(nc)

Numerically solve[6] for -dW = dS = A to get the optimal beds c_0

Calculate $p_0 = \varepsilon/c_0$

Store (c_0, p_0) in a list of ordered pairs

Plot the list of (c_0, p_0) to generate the optimality curve

---

[2] As $\rho < 1$, ${}^{\varepsilon}/_{c} < 1$, so $c > \varepsilon$. We hence choose the minimum number of beds as the smallest integer greater than ε.

[3] This is the list of points that generate our $W(c)$ curve

[4] This is the list of points that generate our $W'(c)$ curve. We drew it to observe its general behaviour.

[5] An exponential approximation was seen to fit the curve well due to analysis of the 1st, 2nd, 3rd, and 4th differences

[6] As we wrote $W(c) = me^{nc}$, then $-W'(c) = -nme^{nc}$. Solving for optimal beds, $-W'(c) = S'(c) = A = -nme^{nc}$. And so $c_o = \frac{1}{n}\ln\frac{-A}{mn}$