A Study of New York City Obstetrics Units Demonstrates the Potential for Reducing Hospital Inpatient Capacity

Linda V. Green, PhD
Armand G. Erpf Professor of the Modern Corporation
Columbia Business School
Columbia University
New York, NY 10027
Phone: 212 854 4108
Email: lvg1@columbia.edu

Nan Liu, PhD
Assistant Professor
Department of Health Policy and Management
Mailman school of Public Health
Columbia University
New York, NY 10032
Phone: 212 304 5591
Fax: 212 305 3405
Email: nl2320@columbia.edu

Key Words. Hospitals, capacity planning, obstetrics, queueing theory

1 Corresponding author
A Study of New York City Obstetrics Units Demonstrates the Potential for Reducing Hospital Inpatient Capacity

Abstract

Hospitals are under significant pressure from payers to reduce costs. The single largest fixed cost for a hospital is inpatient beds, yet there is significant variation in hospital capacity utilization. We study bed capacity in NYC hospital obstetrics units and find that while many hospitals have an insufficient number of beds to provide timely access to care, overall there is significant excess capacity. Our findings, coupled with current demographic and clinical practice trends, indicate that a large fraction of obstetrics units nationwide could likely reduce their bed capacity while assuring timely access to care, resulting in large savings in capital and staffing costs. Given emerging healthcare delivery and payment models that will likely decrease demand for other types of hospital beds, our study suggests that data-based methodologies should be used by hospitals and policy makers to identify opportunities for reducing excess bed capacity in other inpatient units as well.

Key Words: Hospitals, Capacity Planning, Obstetrics, Queueing Theory

Running title: A Study of NYC Obstetrics Units
INTRODUCTION

As a result of cuts from Medicare due to the Patient Protection and Affordable Care Act (PPACA) and increased pressure from private payers to control costs, many hospitals are actively trying to substantially cut their budgets (Palmer, 2013). One obvious candidate for accomplishing this is to reduce the number of beds (Weisman, 2013), as hospital beds represent the largest fixed cost incurred by hospitals (Roberts, et al., 1999). While cutting excess beds makes sense in terms of increasing efficiency and eliminating waste, cutting too many may threaten patient care and actually increase costs by increasing hospital readmissions and internal readmissions to intensive care units (ICUs), reducing quality of care in maternity units and increasing hospital length-of-stay (Chan, Farias, Bambos, & Escobar, 2012; M. Freeman, Savva, & Scholtes, 2013; Kc & Terwiesch, 2012). Therefore, it is important for hospitals to identify and maintain an appropriate level of bed capacity.

There is evidence of significant variation in the utilization of hospital beds. According to the Dartmouth Atlas project, some regions had more than twice the number of acute care beds per capita than other regions even after adjusting for differences in age and sex (Goodman, Fisher, Bronner, Policy, & Practice, 2009). This variation suggests that the investment in acute care beds does not reflect the “true” regional need for hospital acute care.

Hospital beds are not all the same. Different clinical units are used for patients with different health conditions and demand profiles, e.g., cardiac care units, general medical-surgical units, ICUs, pediatric units, psychiatric units and obstetrics units. These units may have very different sizes and differing associated levels of desired access. For example, while some units are used primarily for elective surgical patients who may safely be able to wait days or even weeks for a
bed, others are needed by emergent patients, such as laboring mothers or patients experiencing a stroke or heart attack, and long delays for these beds could result in adverse health outcomes. However, traditional capacity planning guidelines are generally “one size fits all” approaches based on a target bed utilization level and do not consider these size and access factors (Green & Nguyen, 2001). As we demonstrate in this paper, this can result in inappropriate capacity investment suggestions.

In this article, we use hospital obstetrics units, an important and yet understudied setting, as an example to study capacity utilization in hospital inpatient units. Excess obstetrics beds (i.e., postnatal maternity beds) clearly represent a cost saving opportunity for hospitals (Boutsioli, 2010; Gaynor & Anderson, 1995), while insufficient obstetrics capacity creates delays in admitting postnatal mothers as well as laboring mothers because congestion in obstetrics units may back up to labor and delivery areas. These delays may cause adverse health events for both mothers and babies (Cochran & Bharti, 2006; Fetter & Thompson, 1965). Furthermore, a hospital facing insufficient obstetrics capacity may choose to discharge postnatal mothers earlier than desired in order to accommodate new patients, resulting in possible complications, readmissions and patient dissatisfaction (Fink, 2011).

Using a dataset from New York City (NYC), we find that almost none of the NYC hospitals have an appropriate level of obstetrics capacity. Based on our estimates, the overall excess obstetrics capacity in NYC was 184 beds per year during 2008-2009, whereas the total number of insufficient beds citywide amounted to 55 beds per year in the same period, resulting in a net excess of 129 beds per year or approximately 9% of the whole city’s obstetrics capacity. We
estimate that if obstetrics capacity were allocated appropriately in NYC, the financial savings would be substantial, amounting to 26.4 million dollars per year.

We chose to study obstetrics capacity because the beds in obstetrics units are managed independently of the rest of the hospital. Unlike many other inpatient units, the admissions to these beds are restricted to a very specific and well-defined patient population and these patients are not placed in alternative units when an obstetrics bed is not available. Thus, an obstetrics unit is like an isolated “island” in the hospital, and its capacity can be evaluated and planned independently. The common wisdom for planning obstetrics capacity (as in other types of inpatient units) is to set a target for average occupancy levels, i.e., to assure that the average percentage of occupied beds is no larger than a given threshold. The target level for obstetrics units is 75% based on guidelines recommended by the American College of Obstetrics and Gynecology (ACOG) (R. Freeman & Poland, 1997). However, as we will demonstrate, most NYC obstetrics units do not follow this recommendation and, more importantly, achieving this recommended target cannot assure timely access to care simply because that occupancy level does not measure or guarantee the level of access to care received by patients.

Our study uses a data set from the Statewide Planning and Research Cooperative System (SPARCS) of New York State (NYS) which provides patient admission and discharge information at each obstetrics unit. Using this dataset, we recreate the bed census process in each NYC obstetrics unit and evaluate the access to care in each unit based on a target maximum fraction of patients who experience a delay in obtaining a bed. To identify the number of excess/insufficient obstetrics beds in a given hospital, we use a systems modeling methodology, namely queueing theory, to estimate the number of beds needed to provide the desired level of
access to care. Using our empirically derived delay estimates, we show that queueing models generate reliable estimates of desirable levels of bed capacity.

Given the evidence of variation in bed utilization in general, the lack of sensible guidelines for identifying appropriate obstetrics bed capacity, and the likely inertia in adjusting bed capacity as demand patterns change, the variation we find in obstetrics capacity utilization in NYC is likely to be a reflection of that in other parts of the country. Furthermore, excess capacity in obstetrics units is likely to increase in the future if no adjustments are made given the aging population, the concomitant decrease in the birth rate and current efforts to reduce the rate of elective Cesarean sections (C-sections) (Boult, Counsell, Leipzig, & Berenson, 2010; Martin, Hamilton, Ventura, Osterman, & Mathews, 2013; National Institute of Health, 2010). In NYC, for example, we find that though there is a high level of variation in delivery methods across different types of hospitals, a reduction in the overall C-section rate of just 3% (from 33% to 30%) could result in an additional 30 bed savings per year in addition to the 129 bed savings mentioned above. This is due to the longer hospital length-of-stay (LOS) associated with C-section deliveries.

New Contribution

Though there is abundant evidence of variation in utilization of hospital beds (Goodman, et al., 2009), research on how to reduce such variation is relatively scant. Our study demonstrates that approaches based on operational data and systems modeling are more sensible and effective than traditional guidelines in determining appropriate bed capacity levels. Specifically, our study provides insights on reducing inpatient bed capacity based on achieving a desired level of patient access, rather than on arbitrary occupancy level targets. Using a recent dataset of multiple hospitals in NYC, our study is the first to verify the reliability of such data-based approaches.
Our study is also the first to show that reducing C-section rates, even by a small percentage, can lead to large savings in inpatient bed capacity. Our findings, coupled with current demographic and clinical practice trends, indicate that a large fraction of obstetrics units nationwide could likely reduce their bed capacity while assuring timely access to care. Given emerging healthcare delivery and payment models that will likely decrease demand for other types of hospital beds, our study suggests that data-based methodologies should be used by hospitals and policy makers to identify opportunities for reducing excess bed capacity in other inpatient units as well.

CONCEPTUAL FRAMEWORK

Our conceptual framework is based on the premise that timely access to care is the most relevant objective in identifying an “appropriate” level of bed capacity in an obstetrics unit. As discussed earlier, this is in sharp contrast to existing hospital bed capacity guidelines that are based on targeted average occupancy levels. We measure patient access by the likelihood that a patient cannot secure a bed when she needs it (or probability of delay for short) and we estimate this measure using standard queueing models. The authors are not aware of any official or published operational standards regarding patient delays for obstetrics beds, nor any examples of hospitals using queueing theory to estimate obstetrics bed capacity.

METHODS

Data sources

We use a recent SPARCS hospitalization dataset which includes all births in NYC hospitals across the five boroughs from 2008 to 2009 (approximately 120,000 deliveries per year). We include only the 40 hospitals that operated in both 2008 and 2009 in our analysis. The elements in the SPARCS dataset of interest to us include patient admission date and time, discharge date
and time, method of delivery and institution. We are interested in the method of delivery because patients with different delivery methods can have different arrival patterns and lengths of stay (LOS) and thus have different needs for obstetrics bed capacity. We categorize maternal admissions by C-section or vaginal delivery and scheduled or unscheduled admission. According to the data dictionary of SPARCS (2010), a scheduled admission is defined as an admission arranged through the hospital at least 24 hours before the admission. Therefore, there are four categories of admissions: scheduled C-sections, non-scheduled C-sections (emergency C-sections), scheduled vaginal deliveries (inductions) and non-scheduled vaginal deliveries (spontaneous vaginal deliveries).

In order to perform analyses on bed capacities required to meet a given access requirement, we need data on the times of requests for obstetrics beds as well as the LOS for each delivery type. Patient admission and discharge date and time in our dataset correspond to when patients were admitted into and discharged from the hospital, respectively. In most hospitals, a new obstetrics patient first experiences labor and delivery in areas dedicated for these uses before being moved to a postpartum obstetrics bed. So the time at which a bed is requested generally corresponds to the time of the baby’s birth, which is also the time at which the mother’s LOS begins. Unfortunately, these times are not reported in the SPARCS database. So we relied on published data on the durations of labor and delivery for the various delivery alternatives to develop estimates for birth times based on patient admission times. The capacity of each obstetrics unit is measured by the number of staffed beds (not certified or licensed beds) reported in the Institutional Cost Reports (New York State Department of Health, 2010), because staffed beds are beds “in service” and more accurately reflect a unit’s capacity.
Capacity evaluation methods

Our approach is to estimate capacity needs based on the probability of delay experienced by patients in getting a bed. As discussed earlier, the authors are not aware of any official or published standard on the target maximum probability of delay. In an earlier study, Green and Nguyen (2001) conducted an analysis using 1%, 5% and 10% as potential design thresholds for probability of delay to evaluate New York State hospital bed capacity utilization. The 10% standard is the most lenient standard while 1% is the strictest. In our study, we adopt 5% as a “moderate” criterion. That is, we assume that the appropriate level of capacity in an obstetrics unit corresponds to the minimum number of beds that ensures that no more than 5% of the patients will experience a delay in getting a bed on an annual basis. (Due to volume fluctuations over time, the use of this performance metric may result in somewhat lower or higher access levels at various times. We address this issue below.) Therefore, we identify a unit with probability of delay larger than 5% as having insufficient capacity, and as having excess capacity if it can meet this standard with fewer beds.

Our study involves two steps of analyses. The first is to estimate the probability of delay in each NYC obstetrics unit for each study year using the data mentioned above. We do this by using the adjusted in-unit patient admission and discharge times (after accounting for labor and delivery times) to reconstruct the obstetrics patient census process in each hospital using discrete-event simulation (Law & Kelton, 1991). Based on various sources (e.g., Bloom et al. (1998), BabyCenter Medical Advisory Board (2014), Healthline Editorial Team (2012) and Intermountain Healthcare (2014)), vaginal deliveries (including both natural births and induced labor) typically take between 1 and 20 hours while a scheduled C-section normally takes only a
couple of hours. Non-scheduled C-sections often arise after a difficult laboring process, and therefore their durations are close to those of vaginal deliveries.

Based on the information from these sources, we make the following distributional assumptions on delivery times in our simulation: the delivery time for a scheduled C-section is normally distributed with mean 2 hours and standard deviation 20 minutes; for other patient types, the labor and delivery time is normally distributed with mean 10 hours and standard deviation 3 hours. The mean delivery times are chosen according to our findings above, and we set the value for standard deviations so that the range within three standard deviations from the mean generally covers the range of the delivery times reported. To avoid extreme samples (e.g., negative delivery times), we use truncated normal distributions for each type of delivery with sample spaces covering the range of three standard deviations from the mean.

In our simulation, we randomly sample a labor and delivery time for each patient depending on her delivery type. Then the time at which the patient requests an obstetrics bed is calculated as her hospital admission time plus her labor and delivery time. We follow this methodology to sample bed request times for all patients in each year and for each unit. We assume that a patient experiences a delay in bed placement if all staffed obstetrics beds are occupied at the time of a bed request. The empirical estimate for probability of delay in a unit is calculated as the percentage of patients who encounter such a delay. We repeat this sampling process 10 times for each unit and for each year, and report mean probability of delay for each unit (the 95% confidence intervals are very narrow and thus not reported).

The second step of our study is to estimate the minimum number of beds required for each obstetrics unit to achieve the target maximum 5% probability of delay. Our estimation is based
on *queueing theory*, an advanced mathematical modeling technique that is used to predict
customer waiting times in a service system facing periodic congestion due to randomness in its
customer demand and service processes (Gross & Harris, 1985). One advantage of using
queueing models is that they require little data and result in relatively simple formulas for
predicting system performance, thus making it an easy-to-use tool for capacity planning.
Queueing theory can translate the data inputs into a variety of performance measures related to
customer waiting including the probability of delay, the measure of particular interest to us.

We use the classic Markovian queueing model, the M/M/c queueing system, to model obstetrics
units in which postnatal mothers and obstetrics beds act as “customers” and “servers” in the
model, respectively. This choice is driven by both the empirical distributional information of
patient arrivals and adjusted in-unit LOS in our data as well as theoretical results regarding large
queueing systems.

The M/M/c queueing model assumes that 1) customer arrivals follow a Poisson process, where
interarrival times form a sequence of independent and identically distributed (i.i.d.) exponential
random variables, and 2) the service times of customers also constitute a sequence of i.i.d.
exponential random variables (Gross & Harris, 1985). In order to assess the reliability of using
this model, we examined the coefficient of variation (COV), defined as the ratio of the standard
deviation to the mean, for both the inter-arrival and service times for each hospital in our data.
COV measures variability; a larger COV indicates more variability while a zero COV implies
deterministic inter-arrival times (or service times). All else being equal, a queueing system with
less variability experiences less congestion and has a smaller probability of delay for customers.
An exponential distribution has a COV of 1. Most of the 40 units in our study have COVs of
inter-arrival times slightly larger than 1. In particular, the mean unit-specific COV of inter-arrival times is 1.07 in 2008 and the range is [0.93, 1.24]; only two hospitals have COVs of inter-arrival times smaller than 1. These descriptive statistics are respectively 1.07 and [1.00, 1.13] in 2009. For LOS, 37 out of the 40 units have COVs smaller than 1 in 2008, and this count is 38 in 2009. The mean COV across all units in 2008 is 0.69 and the range is [0.33, 1.08]. In 2009, these statistics are 0.70 and [0.35, 1.23], respectively. Given the magnitudes of these COVs and the fact that the smaller variability in the service process is somewhat offset by the relatively larger variability in the arrival process (and, as we discuss later, the variability between weekdays and weekends), we hypothesize that the M/M/c queue is a plausible model to use.

To further support the use of the M/M/c model, we invoke what it is known as the insensitivity property of the M/G/infinity queue. That is, it has been established that as the number of servers in a queueing system grows large, the stationary distribution of number in the system, and hence the probability of delay, becomes insensitive to the distribution of service time and depends only on the mean. Therefore, in practice, the delay probability formula for the M/M/c model is often used to approximate the delay probability for queueing systems with non-exponential service-time distributions, when the number of servers is large, i.e., greater than 10 (Kimura, 1995). All units in our analysis have no fewer than 12 beds and so the use of the M/M/c approximation is reasonable from this perspective as well.

To populate these queueing models, we need only three input parameters for each unit: yearly average patient arrival (admission) rate, yearly average in-unit LOS and the number of staffed beds, all of which are easily calculated from or directly available in our data. To validate the reliability of our method, we use our unit-specific queueing models to estimate the probability of
delay for each unit in each study year based on the historical admissions and LOS data. We then compare these queueing model-based estimates for each unit with the corresponding empirical estimates of the yearly average probability of delay obtained above. We find that the Markovian queueing models generate fairly accurate estimates for probability of delay, particularly at the levels desired for good access to care.

One concern with the use of yearly average demand to estimate capacity needs is that the suggested capacity could underestimate the actual needs during peak demand times, if any. To explore this issue, we first conduct a temporal analysis of patient demand, and then we use simulation to evaluate the probability of delay in each obstetrics unit during peak times assuming that the unit adopts the capacity suggested by our queueing model based on the yearly average demand. We find that none of units are highly congested even during the peak times, validating that our approach is a reliable capacity planning method. Details on this validation are discussed in the next section.

Using these queueing models, we also evaluate the impact on capacity of evolving changes in practice, namely, reduction in the C-section rate. Because hospital LOS is generally much longer for patients who have a C-section, a decrease in the C-section rate will reduce the mean LOS in a unit and therefore the number of beds needed to ensure a given maximum probability of delay.

RESULTS

The average capacity of the 40 NYC obstetrics units in 2008-2009 was 34.5 beds with range [12, 80] and a total capacity of 1381 beds. These hospitals collectively handled around 240,000
maternal admissions in 2008-09. Table 1 shows this information in detail and patient LOS statistics by type. The overall C-section rate in NYC during 2008-2009 was about 33%, close to the national level of 2007 (Menacker & Hamilton, 2010).

[Insert Table 1 here]

Variation of capacity utilization in NYC obstetrics units

The reconstructed census from each NYC obstetric unit reveals a high degree of variation in bed capacity utilization across these units. Figure 1 shows the census processes of two sample units in 2009. The census of the unit on the left rarely hit 32 though this unit had 42 staffed beds. In the unit on the right, however, the unit was often completely full resulting in about one fifth of the postpartum mothers being delayed in bed placement.

We estimate that the least congested unit had a yearly occupancy level of 34%, while the busiest one was apparently full virtually all of the time. Though, as mentioned previously, the suggested average occupancy level by ACOG is 75%, fewer than 25% of these units had an occupancy level in the range of 70-80%. It is also important to note that this suggested 75% guideline does not assure timely access to care, as demonstrated by the unit on the right of Figure 1 which had an observed average occupancy of 73%. This is because smaller units, such as this one with only 14 beds, must operate at lower occupancy levels in order to achieve high levels of access.

[Insert Figure 1 here]

If we use the probability of delay to evaluate capacity utilization, we find that 8 of the 40 NYC obstetrics units in 2008 were significantly underutilized with probability of delay less than 0.1%, while 11 were overutilized with probability of delay larger than 10%. These two counts are 9
and 12 in 2009, respectively. Interestingly, four units adjusted their capacity during 2008-09 (see Table 2) but none of the underutilized or overutilized units made capacity adjustment. Among the four units which adjusted their capacities, only two (hospitals C and D) made an adjustment consistent with bringing the probability of delay closer to 5%.

[Insert Table 2 here]

**Validation of queueing models**

By comparing the empirical estimates of probability delay in each unit obtained from the discrete-event simulation technique with those estimated by the queueing models, we find that for obstetrics units with empirical yearly average probability of delay less than 5%, the corresponding queueing model identified these perfectly in both years. That is, queueing models also predict that in these units the average probability of delay is less than 5%. For units with yearly average delay probability no smaller than 5%, the queueing models successfully identified these more congested units in 82% cases (14/17) in 2008 and in 87% cases (13/15) in 2009. In the five cases where the queueing model misclassified these units, four of them have the empirically estimated yearly average probabilities of delay less than 8% and one less than 12% indicating that they were not very congested.

**Assessment of current capacity utilization**

We now use the M/M/c queueing model to estimate the appropriate capacity level for each unit using 5% as a performance threshold for probability of delay. Table 3 compares these appropriate levels to the actual numbers of obstetrics beds. Based on this analysis, the vast majority of the hospitals (more than 37 out of 40) did not have appropriate levels of bed capacity. The total estimated number of excess beds citywide was more than 183 each year, while the
overall deficit of capacity amounts to more than 54 beds each year. More importantly, the capacity in many of these units significantly deviates from the appropriate level. Almost half of the hospitals had 10% or more excess capacity, while one fifth of the hospitals had a deficit of beds that is at least 10%.

[Insert Table 3 here]

To further analyze this variation, we examine capacity utilization by type of hospital. We follow the definition of hospital type provided by the United Hospital Fund (UHF), a large NYC-based nonprofit health services research and philanthropic organization: academic medical centers (AMC); Health and Hospital Corporation hospitals (HHC) which are public; private hospitals where Medicaid and self-paid patients account for over 50% of discharges (High Mcaid); and other private hospitals (Others). HHC and High Mcaid hospitals usually provide care to underserved populations with relatively low socioeconomic status. We find that inappropriate obstetrics capacity is found in every type of hospital, and in particular, High Mcaid hospitals seem to be more likely to have excess capacity compared to others.

[Insert Figure 2 here]

To check if our suggested bed capacities based on yearly average demand are appropriate given temporal patterns of demand, we examined weekly demand fluctuations to determine if any units would have been overly congested at these levels during peak demand times. We only considered the day-of-week effect and do not examine the potential impact of monthly seasonality for two reasons. First, though our data revealed higher demand levels during the summer months, hospitals are unlikely to plan their regular bed capacity based on a couple of peak demand months. Rather, they are likely to employ other strategies, such as using “swing”
beds as flexible capacity. Second, our two years of data are not sufficient to estimate the true level of this seasonality.

During our study period, daily patient demand was on average 43% higher on weekdays than on weekend days. Breaking it down by delivery type, daily weekday demand was 160%, 70%, 40% and 20% more than daily weekend demand for scheduled C-section, nonscheduled C-sections, scheduled vaginal deliveries and nonscheduled vaginal deliveries, respectively. As expected, substantially more scheduled C-sections were performed during weekdays than weekends.

Based on our simulation using the capacities suggested by the queueing models, we estimate that the citywide weighted average probability of delay for weekday patients would have been 8.2% in 2008 and 8.0% in 2009, with weights being proportional to the total patient admissions to each hospital. In 2008, 34 out of 40 units would have had an estimated mean probability of delay less than 10%, 5 units between 10% to 14% and 1 unit at 18%. In the last unit, 83% of the deliveries are scheduled ones, which mainly occur during weekdays, explaining its relatively high level of delays in weekdays. In 2009, all units would have had an estimated mean probability of delay less than 14%, and 33 units less than 10%. Thus, the vast majority of obstetrics units would not have been excessively congested during these peak times using the queueing-based capacities based on annual daily demand. If one wanted to keep average weekday delays below 10% for all hospitals, this could be achieved with one additional bed at each hospital not meeting this standard for a total of 7 additional beds.

**Variation in types of delivery**

Our study also identifies very high levels of variation with respect to types of delivery. The two-year average rates of scheduled C-section, unscheduled C-section, scheduled vaginal delivery
and unscheduled vaginal delivery in these 40 units vary in the ranges of [0.0%, 38.6%], [0.1%, 39.8%], [0.0%, 66.8%] and [0.1%, 79.1%], respectively. The aggregate rate of scheduled delivery (scheduled C-section and scheduled vaginal delivery combined) varies in [0.0%, 99.8%] among these units, while the overall C-section rate (scheduled plus non-scheduled) ranges in [20.9%, 46.3%]. To analyze this variation, we investigate the distribution of delivery types by borough and by type of hospital (see Figure 3). We find that HHC hospitals and High Mcaid hospitals have much lower rates of scheduled delivery and higher rates of non-scheduled C-section compared to other hospitals. In contrast, Manhattan has the highest rate of scheduled deliveries compared to the other boroughs.

[Insert Figure 3 here]

**Impact of reducing C-sections**

Finally, we evaluate the impact of reducing the C-section rate on the bed capacity needed to assure the delay probability to be less than 5% in each unit. We assume that the C-section rates in all units with “high” levels are reduced to some common “appropriate” level. However, there is no consensus on what an appropriate C-section rate should be for scheduled or nonscheduled C-sections in the US. In the global context, the World Health Organization (1985) recommends an overall rate between 10 to 15%. Since the C-section rate in the U.S. has never been below 15% (Menacker & Hamilton, 2010), using 15% as the appropriate target level may not be realistic.

To obtain greater insight into what might be an appropriate target level for C-section rates, we considered medical risks. For low-risk patients, the American College of Obstetricians and Gynecologists (2000) sets a target rate of 15.5 percent. For high-risk populations, there is
evidence that the C-section rate can be controlled at around 20%. For example, Geisinger Health System was able to reduce its C-section rate to about 23% after implementing evidence-based guidelines, even though they serve a generally poorer and high-risk population (Berry, et al., 2011). San Francisco General Hospital, which is largely a hospital for the poor, maintains a C-section rate of 20.67% following evidence-based medicine (Rosenberg, 2014).

The above suggests that it may be reasonable to assume that an overall C-section rate of about 20% is achievable in NYC. However, to obtain a broader picture of the potential impact of reduced C-sections on beds savings, we consider “target” C-section rates of 25% and 30% as well. Though we do not presume that any specific level would be achievable at every individual hospital, to get a ballpark estimate of the impact of reducing C-section rates citywide, we assume that the specified target level is met for each hospital. More specifically, for units with C-section levels higher than the target, we keep the ratio of unscheduled and scheduled C-sections the same and proportionally reduce their rates so the overall rate achieves the target. For units with C-section levels lower than the standard used, we assume no changes.

Table 4 compares both the actual bed capacity at each unit and the estimated levels assuming a maximum probability of delay of 5% and that the C-section rates were to be reduced to the target level. As we showed earlier, if the capacities of NYC obstetrics units were set at appropriate levels as estimated by queueing models, then the city could have safely closed 129 beds in 2008 (see the net excess in Table 3). If, in addition, the overall C-section rate were reduced to 30% (i.e., a 3% reduction from its current level of 33%), then an additional 30 beds could be saved based on 2008 data (see the increment of the net excess bed capacity from Table 4 to Table 3). Similarly, if the overall NYC C-section rate could be reduced to 25% (or 20%), additional cuts
could be made of 54 (or 85) beds citywide. These bed savings are similar in 2009. Roughly speaking, for every one percent reduction of C-sections in NYC, the city could save approximately 7 beds citywide per year. Furthermore, these savings do not reflect the impact of the reduced variability in daily demand due to fewer scheduled deliveries, see discussion section below.

[Insert Table 4 here]

DISCUSSION AND CONCLUSIONS

Using a recent dataset of NYC obstetrics units, our study reveals significant variation across NYC obstetrics units in their capacity utilization; almost none of these units have matched their capacity and patient demand effectively. Furthermore, such mismatch is found in all types of hospitals (Figure 2), suggesting that the capacity utilization problem is likely ubiquitous. A closer look at this variation indicates that the overall excess obstetrics capacity in NYC is more than 183 beds (Table 3). Though the estimated cost of an unused bed varies across studies, ranging from $80,951 (Gaynor & Anderson, 1995), $103,150 (Institute of Medicine and National Academy of Sciences, 1976), to $205,105 (Keeler & Ying, 1996), these estimates all point to a significant amount of wasted resources due to excess hospital bed capacity (these figures have been converted to 2014 dollars using a 3% annual inflation rate). Conversely, many obstetrics units in NYC lack enough capacity to provide timely care to maternal patients. The overall shortage is more than 54 beds citywide (Table 3). Insufficient bed capacity could lead to congestion in the delivery units, placement of postpartum mothers into inappropriate areas and earlier discharge of patients than desired, all of which might be associated with adverse patient outcomes and result in additional healthcare costs (Griffin, Xia, Peng, & Keskinocak, 2012). In
summary, the net excess obstetrics capacity in NYC is approximately 129 beds per year and would be only a bit less even if we account for a few additional beds to deal with higher demands during weekdays. If obstetrics beds were allocated appropriately citywide, the financial savings due to this reallocation alone could amount to between 10.4 to 26.5 million dollars per year based on the cost estimates above. This financial saving may be conservative as it does not account for potential savings due to improved patient outcomes and reduced complications as a result of improved patient access to care.

The variation observed above, and, in particular, excess capacity, is likely to persist and perhaps worsen if no proper adjustments are made due to the ongoing demographic changes in the US. The US population is aging quickly and women are delaying having children and having fewer of them, driving the national birth rate to continuously decrease (Boult, et al., 2010; Martin, et al., 2013). In addition, as demonstrated by our findings on variation in delivery type, there is substantial evidence of overuse of Cesarean sections and inductions. As a result, efforts are underway to address this problem to both achieve better clinical outcomes and cost savings (Kozhimannil, Law, & Virnig, 2013). For instance, an expert panel convened by the National Institute of Health in March 2010 systematically reassessed different delivery methods after a prior Cesarean, and urged steps to mitigate current barriers to Vaginal Birth after Cesareans (National Institute of Health, 2010). Many health care systems are also engaged in efforts to reduce such overuse. In addition to the two examples discussed earlier, Intermountain Healthcare was able to reduce its rate of elective inductions down to 2% and the rate of overall C-sections to 21% following evidence-based guidelines and quality improvement practices (James & Savitz, 2011).
Our findings from NYC complement these efforts by suggesting that the cause for variation in delivery method has a significant socio-demographic component. HHC and High Mcaid hospitals have much lower rates of scheduled delivery and much higher rates of non-scheduled C-section compared to others. This may reflect a lack of necessary prenatal care for the population served by these hospitals (Metzger, et al., 2007). On the other hand, Manhattan has the highest rate of scheduled deliveries across all boroughs. Considering that Manhattan is the most affluent borough in NYC with the highest number of working professionals and likely fewer high-risk patients, this suggests that a high level of scheduled deliveries is driven by patient and physician choices and convenience, rather than clinical appropriateness. This is consistent with the finding from a recent national level study which reveals a higher-than-average variation in C-section rates among low-risk patients (Kozhimannil, et al., 2013).

While earlier studies focus on savings in variable costs due to a decrease in unnecessary C-sections, our study demonstrates that cost savings can be far greater if hospitals also adjust their capacity appropriately to achieve a reduction in fixed costs. Using NYC as an example, we find that if hospitals could reduce the overall C-section rate to 20%, a level comparable to what Geisinger, San Francisco General or Intermountain has achieved, approximately 215 postpartum beds or 16% of the total obstetrics capacity per year could be eliminated while assuring timely access to care. This bed reduction may actually be greater because reducing the overuse of elective C-sections, which are mostly performed on weekdays, would “smooth” patient demand over the week so that the average bed delay would be lower than that predicted by a model that assumes the same average demand each day thus enabling the same desired level of access with fewer beds (Whitt, 1980).
In conclusion, our study indicates that in addition to establishing more evidence-based standards for the use of elective delivery methods and providing more prenatal care for underserved populations, hospitals should be using data-based models to identify the number of beds needed to assure timely access for their patients. In addition, guidelines issued by professional organizations as well as governments should be based on the use of such patient access oriented analyses rather than traditional target occupancy levels.

Specifically, our study provides support for the use of queueing models to plan bed capacity, echoing the recommendations of a recent Institute of Medicine report issued in 2011 (Grossmann, Goolsby, Olsen, & McGinnis, 2011). Though we have demonstrated its use and reliability for obstetrics units, we argue that these types of models can and should be used for other inpatient units as well. Similar methods have been used previously to investigate hospital capacity planning issues (Green, 2006), and the increasing availability of operational hospital data makes it easier to use such models.

In an environment in which it has become critical to reduce unnecessary costs, it is imperative to identify excess inpatient bed capacity. Given the variation in hospital utilization documented by the Dartmouth Atlas, and the introduction of new Medicare policies aimed at reducing admissions and readmissions through the use of penalties, bundled payment schemes and the establishment of Accountable Care Organizations (ACOs), ample evidence exists that there is significant excess bed capacity in many hospitals which could be exacerbated by decreasing demand for inpatient care. As in other industries which have used this data-based methodology for decades (Gilliam, 1979; Koole & Mandelbaum, 2002; Newell, 1982), hospitals need to start...
adjusting capacities based on scientific methods aimed at assuring timely access and containing
cost rather than on discredited target occupancy levels or guesswork.
REFERENCES


Palmer, K. (2013). Cleveland Clinic announces job cuts to prepare for Obamacare. *Reuters (Sep 18)*.


<table>
<thead>
<tr>
<th></th>
<th>Scheduled C</th>
<th>Non-scheduled C</th>
<th>Scheduled V</th>
<th>Ordinary V</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of admissions</td>
<td>31338</td>
<td>48612</td>
<td>51107</td>
<td>111136</td>
<td>242193</td>
</tr>
<tr>
<td>Percentage</td>
<td>12.9%</td>
<td>20.0%</td>
<td>21.1%</td>
<td>45.9%</td>
<td>100%</td>
</tr>
<tr>
<td>Mean LOS (days)</td>
<td>4.12</td>
<td>3.85</td>
<td>2.14</td>
<td>2.20</td>
<td>2.76</td>
</tr>
<tr>
<td>Stdev LOS (days)</td>
<td>2.85</td>
<td>3.12</td>
<td>1.26</td>
<td>1.41</td>
<td>2.23</td>
</tr>
<tr>
<td>COV LOS</td>
<td>0.69</td>
<td>0.81</td>
<td>0.59</td>
<td>0.64</td>
<td>0.81</td>
</tr>
</tbody>
</table>
Figure 1 Bed census process from two sample obstetrics units
Table 2 Changes of beds in obstetrics units from 2008-09

<table>
<thead>
<tr>
<th>Hospital</th>
<th>Mean estimated probability of delay in 2008</th>
<th>Mean estimated probability of delay in 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hospital A</td>
<td>2.0%</td>
<td>1.1%</td>
</tr>
<tr>
<td>Hospital B</td>
<td>6.4%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Hospital C</td>
<td>0.2%</td>
<td>3.2%</td>
</tr>
<tr>
<td>Hospital D</td>
<td>8.5%</td>
<td>4.9%</td>
</tr>
</tbody>
</table>

Note: these percentages are mean probability of delay estimated by the simulation program.
Table 3 Comparison of appropriate capacity levels with current levels in NYC obstetrics units

<table>
<thead>
<tr>
<th></th>
<th>2008</th>
<th>2009</th>
</tr>
</thead>
<tbody>
<tr>
<td># of hospitals with excess capacity</td>
<td>24</td>
<td>24</td>
</tr>
<tr>
<td># of hospitals with &gt;10% excess capacity</td>
<td>19</td>
<td>17</td>
</tr>
<tr>
<td>Mean excess (beds/unit)</td>
<td>7.6</td>
<td>7.7</td>
</tr>
<tr>
<td>Median excess (beds)</td>
<td>7.0</td>
<td>6.0</td>
</tr>
<tr>
<td>Total excess (beds)</td>
<td>183</td>
<td>184</td>
</tr>
<tr>
<td># of hospitals with insufficient capacity</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td># of hospitals with &gt;10% insufficient capacity</td>
<td>7</td>
<td>8</td>
</tr>
<tr>
<td>Mean slack (beds/unit)</td>
<td>3.9</td>
<td>4.2</td>
</tr>
<tr>
<td>Median slack (beds)</td>
<td>2.5</td>
<td>4.0</td>
</tr>
<tr>
<td>Total slack (beds)</td>
<td>54</td>
<td>55</td>
</tr>
<tr>
<td><strong>Net excess</strong></td>
<td>129</td>
<td>129</td>
</tr>
</tbody>
</table>
Figure 2 Capacity Utilization of Obstetrics Units by Type of Hospitals
Figure 3 Variation in types of delivery by boroughs and by hospital types
**Table 4 Impact of reducing C-section rate on obstetrics capacity needs**

<table>
<thead>
<tr>
<th></th>
<th>2008(^a)</th>
<th>2009(^a)</th>
<th>2008(^b)</th>
<th>2009(^b)</th>
<th>2008(^c)</th>
<th>2009(^c)</th>
</tr>
</thead>
<tbody>
<tr>
<td># of hospitals with excess capacity</td>
<td>28</td>
<td>28</td>
<td>26</td>
<td>28</td>
<td>27</td>
<td>24</td>
</tr>
<tr>
<td># of hospitals with insufficient capacity</td>
<td>10</td>
<td>12</td>
<td>12</td>
<td>9</td>
<td>11</td>
<td>12</td>
</tr>
<tr>
<td>Total # of excess beds</td>
<td>244</td>
<td>221</td>
<td>203</td>
<td>247</td>
<td>223</td>
<td>204</td>
</tr>
<tr>
<td>Total # of insufficient beds</td>
<td>30</td>
<td>38</td>
<td>44</td>
<td>29</td>
<td>40</td>
<td>46</td>
</tr>
<tr>
<td><strong>Net excess in beds</strong></td>
<td><strong>214</strong></td>
<td><strong>183</strong></td>
<td><strong>159</strong></td>
<td><strong>218</strong></td>
<td><strong>183</strong></td>
<td><strong>158</strong></td>
</tr>
</tbody>
</table>

\(^a\): assuming overall C-section rate reduced to 20% at all hospitals whose C-section rate > 20%

\(^b\): assuming overall C-section rate reduced to 25% at all hospitals whose C-section rate > 25%

\(^c\): assuming overall C-section rate reduced to 30% at all hospitals whose C-section rate > 30%