Tactical Increases in Operating Room Block Time Based on Financial Data and Market Growth Estimates from Data Envelopment Analysis

Liam O’Neill, PhD*
Franklin Dexter, MD, PhD†

BACKGROUND: Data envelopment analysis (DEA) is an established technique that hospitals and anesthesia groups can use to understand their potential to grow different specialties of inpatient surgery. Often related decisions such as recruitment of new physicians are made promptly. A practical challenge in using DEA in practice for this application has been the time to obtain access to and preprocess discharge data from states.

METHODS: A case study is presented to show how results of DEA are linked to financial analysis for purposes of deciding which surgical specialties should be provided more resources and institutional support, including the allocation of additional operating room (OR) block time on a tactical (1 yr) time course. State discharge abstract databases were used to study how to perform and present the DEA using data from websites of the United States’ (US) Healthcare Cost and Utilization Project (HCUPNet) and Census Bureau (American FactFinder).

RESULTS: DEA was performed without state discharge data by using census data with federal surgical rates adjusted for age and gender. Validity was assessed based on multiple criteria, including: satisfaction of statistical assumptions, face validity of results for hospitals, differentiation between efficient and inefficient hospitals on other measures of how much surgery is done, and correlation of estimates of each hospital’s potential to grow the workload of each of eight specialties with estimates obtained using unrelated statistical methods.

CONCLUSIONS: A hospital can choose specialties to target for expanded OR capacity based on its financial data, its caseloads for specific specialties, the caseloads from hospitals previously examined, and surgical rates from federal census data.


Data envelopment analysis (DEA) is a technique that hospitals and anesthesia groups can use to understand their potential to grow different specialties of inpatient surgery (Table 1). Rather than recruit another vascular surgeon and then wait to see if workload increases, a hospital can determine beforehand whether there is potential for market growth in vascular surgery. Previous papers have described the methodology, focusing sequentially on establishing: validity (1), usefulness for several applications (2), and presentation of results (3). Those three papers describe how to benchmark each specialty’s surgical caseload at a study hospital relative to caseloads at all other hospitals in the study hospital’s state.

In the current paper, we consider a hospital at which surgical workload had increased progressively over 7 yr. However, during the next 7 yr, inpatient surgical workload had not increased. Outpatient surgery had declined, as competing facilities were opened by physicians. Efforts to maintain workload included the allocation of large amounts of operating room (OR) block time. The hope had been to have sufficient excess of revenues to variable costs (i.e., contribution margin) to offset the large capital and fixed staffing costs. These efforts were failing, resulting in a more precarious financial condition. The question asked by the administrators and anesthesiologists was whether there were specialties to target for short-term institutional support to promote growth. Such surgical specialties would need to have both market growth potential and a relatively high contribution margin.
Table 1. Description of the Data Envelopment Analysis Model

<table>
<thead>
<tr>
<th>Outputs—Numbers of hospital discharges including the listed procedures, obtained from the corresponding CCS or DRG (1)</th>
<th>Specialty for which listed procedures are a reliable surrogate (1,2) for the ‘inpatient workload’</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA</td>
<td>Abdominal aortic aneurysm resection</td>
</tr>
<tr>
<td>CABG</td>
<td>Coronary artery bypass graft</td>
</tr>
<tr>
<td>Colorectal</td>
<td>Colorectal resection</td>
</tr>
<tr>
<td>Craniotomy</td>
<td>Craniotomy, not for trauma</td>
</tr>
<tr>
<td>Hip replacement</td>
<td>Hip replacement</td>
</tr>
<tr>
<td>Hysterectomy</td>
<td>Hysterectomy</td>
</tr>
<tr>
<td>Lung resection</td>
<td>Lung resection</td>
</tr>
<tr>
<td>Nephrectomy</td>
<td>Nephrectomy</td>
</tr>
</tbody>
</table>

Inputs

| Beds | Staffed acute and intensive care beds at the hospital |
| County | Estimated total hospital charges for the above eight procedures performed on residents of hospitals |
| Region | County and region, normalized by the land square miles of hospital’s county and region |
| Surgeons | Surgeons who performed at least three cases of any one of the above eight procedures at the hospital |
| Tech | Number of nine high technology services (2) offered at the hospital (e.g., solid organ transplantation) |

The limiting factor to answering such a question scientifically is neither the financial analysis nor the DEA. Every hospital that bills Medicare and knows its implant costs has the needed financial data (4). Performing the financial analysis in Excel takes 1–2 h with appropriate software. Performing the DEA itself takes about 30 min.

The challenge has been the time and effort required to obtain and process a new set of state discharge abstract data to obtain an abstract of the hospitalization of each patient undergoing one of the studied procedures at every hospital in the state. First, approval is obtained from the hospital’s state for the data, which (appropriately) involves multiple forms. Second, the data need to be cleaned, as the quality of fields varies among states. These time-consuming steps can make the DEA seem impractical.

Instead of using state discharge abstract data to measure the actual rates at which specific procedures were performed, the analysis could potentially be based on estimates derived from readily-available national data. For example, the age- and gender-adjusted rate of lung lobectomy in the United States (US) can be obtained in about 10 min from the website of the US Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project’s data warehouse (http://hcup.ahrq.gov/) (5). The same applies to the other common procedures listed in Table 1. In addition, the population by age and gender can be determined for all US counties from the Census Bureau’s website (http://factfinder.census.gov/). Multiplying the rate times the population might be sufficient to determine how much surgery is likely being done on residents living near a hospital.

In this paper, we develop such a process that relies on data available from Federal websites. We apply it to the study hospital and its financial data. We use six different criteria to validate the process, and in the process further explore the underpinnings of using DEA to guide selective increases in allocations of OR block time.

CASE STUDY

The case study is presented here in lieu of a separate Background section. The Validation section follows, as it relies on an understanding of the DEA material in the Case Study section.

None of the steps in this case study is new, other than the automatic identification of surgeon specialty, later. We list the specific steps taken for the specific hospital studied. Readers who would be performing the financial analyses and/or the DEA will need to refer to the referenced papers for step-by-step methodologies and equations.

Figure 1 shows the overall contribution margin per OR hour for each surgeon from the study hospital over 1yr, along with the 95% confidence intervals (6). The patients included underwent outpatient surgery or were admitted on the day of their elective surgery (4,6,10). Among these 11,496 cases, 6.2% were of a procedure listed in Table 1. Only elective surgery was studied, because a hospital can alter its elective workload through tactical decisions, while changes in urgent surgery involve longer-term strategic planning.

Figure 2 shows the successive exclusion of specialties for consideration of targeted growth through increases in OR allocations. The steps were described previously (10). Each surgeon is used as a surrogate for his or her specialty. Only increases in OR allocations are considered, because reductions rarely are mandated via tactical mechanisms. If a decision were
made to reduce a specific program, the hospital likely would allow the number of subspecialty surgeons to decrease through attrition. As the subspecialty’s OR workload declines gradually, progressively less OR time would be allocated to reduce expected underutilized OR time (11).

First, surgeons (specialties) with below average contribution margin per OR hour were excluded from consideration of being allocated more block time than needed for their existing cases (10). The reason is that, instead, some or all of the additional OR time could be used as general purpose overflow for specialties that have filled their allocated OR time (12,13) and have more cases to perform (14,15). Both surgeons with below and above average contribution margin per OR hour would have access to that overflow OR time. Thus, the natural expansion of specialties into the overflow time would result in the cases scheduled into overflow OR time having approximately the average contribution margin per OR hour. If much less extra OR time is available versus the expected growth in use of OR time by surgeons (specialties) with high contribution margins per OR hour, some surgeons with above average contribution margins may also be excluded from additional allocations. The Appendix of Ref. (10) gives the equations for calculating the threshold value to use.

Second, among the remaining surgeons (specialties), those with <2 h a week of cases were excluded from further consideration (10).

Third, the hospital administrators’ and anesthesiologists’ judgment was that virtually all of the outpatient cases would be lost for those surgeons who had invested in a physician-owned ambulatory surgery center. For all of these surgeons, once their outpatient cases were excluded, their remaining cases represented <2 h per week of cases. Thus those surgeons (specialties) were excluded (10).

Fourth, intensive care unit (ICU), postanesthesia care unit, and hospital ward beds were considered. There were no reports of cancellations due to limited capacity. Thus, these were not considered constraints in the analysis, which differs from the analyses published for two other hospitals (10,16). In addition, there were no reports of variability among days of the week in the numbers of admissions to ICUs or wards after elective (scheduled) surgery causing disruption of the admission of nonsurgical patients from emergency departments (17). Therefore, variability among days of the week also was not considered, unlike analyses published for other hospitals (18,19).

Fifth, those surgeons (specialties) with wide confidence intervals in estimated contribution margin per OR hour from Figure 1 were identified in Figure 2 (6). The confidence intervals are useful to identify surgeons (specialties) for which outlier patients could cause spurious tactical decisions (9). All such surgeons had been excluded from previous steps. Thus, the uncertainty was not considered in the analysis (9).

The functional specialties of the remaining surgeons in Figure 2 were identified. Each outpatient visit or hospitalization with surgery had a primary International Classifications of Diseases Clinical Modification 9 procedure code (ICD-9-CM). The corresponding Clinical Classifications Software (CCS) code was obtained for each ICD-9-CM from the federal website listed in the legend of Table 1. Thus, there was one CCS for each case. The most common CCS code was obtained for each surgeon. Specialty was determined based on the most common CCS. There was no ambiguity for any surgeon. For example, the two cardiac surgeons in Figure 2 both had “Coronary artery bypass graft” as the most common CCS; the three general surgeons had “Cholecystectomy and common duct exploration;” the five gynecologists had “Hysterectomy, abdominal and vaginal;” and the four urologists had “Procedures on the urethra.”

The next step in the process of allocating increases in OR block times while considering the financial implications was to estimate potential for growth in OR workload (10).

Figure 3 shows a graphical representation of a Productivity Ratio calculated by the DEA. The variables are listed in Table 1. Refs. (3) and (20) give the corresponding equations. The values along the axes
are the annual numbers of Colorectal discharges and staffed acute care hospital Beds.

Hospital discharges with a Colorectal CCS procedure are plotted, because the discharges are proportional to the total workload of general surgery (1,2). The other seven surgical procedures used to represent their specialties are given in Table 1. Appendix 1 of Ref. (1) reviews other potential procedures and why they were not used.

Beds are plotted, because they are a surrogate for the size and the market visibility of the hospital, as relevant for marketing to potential patients (21). Beds were defined as the number of staffed acute care beds as reported in the annual survey of the American Hospital Association. Neither surgical beds nor ORs were used, because it is the number of acute care beds that predicts the market visibility of a hospital (21).

The other two surrogates used for market visibility are the number of Surgeons and high-Technology services (Table 1). Hospitals with larger market visibility generally perform more surgery.

The filled square in Figure 3 shows the study hospital. Open circles are used to show Colorectal discharges and Beds for hospitals in Iowa and Pennsylvania. The study hospital sent its data of Table 1, excluding the County and Region variables. These latter variables are the focus of the Validation section, later. They are estimates of the amount of surgery performed at any hospital on residents of the county and region.

DEA selects automatically those outputs and inputs in Table 1 for the study hospital to appear as efficient as possible at producing surgery. A virtual, benchmark hospital is created simultaneously by DEA and shown as a triangle in Figure 3. The productivity ratio for Colorectal discharges to Beds is the ratio of the slopes of the two lines in Figure 3. Because the line for the study hospital is above the line for its benchmark hospital, the study hospital is performing more general surgery than expected, based on its physical size and visibility.

Table 2 shows the productivity ratios for all combinations of outputs from Table 1 to inputs from Table 1 expressed as percentage differences from the corresponding benchmark hospital. For example, from Figure 3, the Colorectal surgery per Beds of the study hospital was approximately 20% higher than that of the benchmark hospital. The value of the productivity ratio in the corresponding column and row of Table 2 is 20%.

Table 2 shows that CABG and Hysterectomy were the outputs chosen automatically by DEA to make the study hospital appear as efficient as possible. The inputs selected by DEA were Beds, County, Region, and Technology. The combinations of these selected outputs and inputs are listed in bold italic. The efficiency score of the hospital equals approximately 2.0, where 2.0 = 200% = 100% plus the listed bold italic values. The
study hospital produced 2.0 as much surgery as the best other hospitals could if they had the same conditions (inputs).

Because DEA selected which outputs and inputs to compare based on making the study hospital appear as efficient as possible, the efficiency score of 2.0 represents a best-case scenario. The result is based on the premise that the managers have been focusing extra resources on those outputs that would make the hospital as efficient as possible at producing surgery. For example, more OR block time may have been allocated tactically to those specialties that the managers forecasted could grow and use it. Letting DEA selectively choose outputs to study makes sense, as the analysis then reflects the ability of a hospital to focus on the specialties for which the hospital can produce a lot of surgery. A specialty orthopedic hospital may be very efficient, yet will perform poorly at producing thoracic surgery.

The selection of inputs by DEA identifies the bottlenecks to doing more surgery at the hospital. Ref. (2) presents these concepts graphically. However, the graphs are limited to one input and one or two outputs, whereas the productivity ratios of Table 2 show multiple dimensions simultaneously. The equations for the super-efficient DEA model that were used to create Table 2 are given in Refs. (2) and (3).

Some of the values in Table 2 are negative. The study hospital had relatively weak market capture for these outputs: AAA for vascular surgery, Craniotomy for neurological surgery, and Hip Replacement for orthopedic surgery. Readers interested in how the table was produced using DEA should refer to Ref. (3).

Table 3 lists the four hospitals contributing to the virtual, benchmark hospital. The same weight from the second column of Table 3 is applied to all procedures (third through 10th columns) at each of the hospitals (listed as A–D). The selection of the hospitals contributing to the benchmark and their weights are a direct consequence of the selection of outputs and inputs to make the study hospital appear as efficient as possible (2).

A gap is the number of additional cases of each type of procedure that the hospital can be producing based on the hospital’s performance relative to its benchmark hospital. Abbreviations for the outputs (columns) and inputs (rows) are given in Table 1. For example, Surgeons refers to the total number of surgeons at each hospital as is relevant to market visibility, not the number of surgeons for each of the columns in this Table 2 (1,2). Data envelopment analysis chooses automatically outputs and inputs to make the study hospital appear as efficient as possible. The combinations of these selected outputs and inputs are listed in bold italic. The equations for the calculations used to produce the Table are given in Ref. (3). The so-called super-efficient data envelopment analysis model was used (1,3). The software used for the linear programming was the Premium version of Solver, the add-in from Frontline Systems (Incline Village, NV) that came with Microsoft Office Excel 2003 (Redmond, WA). The productivity ratios were rounded to the nearest 10% to make the interpretation easier.

Table 3. Gaps in the Number of Cases of Each Procedure Type at the Study Hospital

<table>
<thead>
<tr>
<th></th>
<th>Weight</th>
<th>AAA</th>
<th>CABG</th>
<th>Colorectal</th>
<th>Craniotomy</th>
<th>Hip replacement</th>
<th>Hysterectomy</th>
<th>Lung resection</th>
<th>Nephrectomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Study Hospital</td>
<td>14</td>
<td>756</td>
<td>186</td>
<td>32</td>
<td>179</td>
<td>459</td>
<td>47</td>
<td>44</td>
<td></td>
</tr>
<tr>
<td>Hospital A</td>
<td>0.64</td>
<td>31</td>
<td>143</td>
<td>57</td>
<td>263</td>
<td>210</td>
<td>26</td>
<td>23</td>
<td></td>
</tr>
<tr>
<td>Hospital B</td>
<td>0.07</td>
<td>52</td>
<td>120</td>
<td>35</td>
<td>147</td>
<td>171</td>
<td>48</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>Hospital C</td>
<td>0.13</td>
<td>53</td>
<td>156</td>
<td>62</td>
<td>160</td>
<td>276</td>
<td>31</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>Hospital D</td>
<td>0.20</td>
<td>85</td>
<td>156</td>
<td>115</td>
<td>274</td>
<td>235</td>
<td>62</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Benchmark</td>
<td>0.48</td>
<td>152</td>
<td>70</td>
<td>235</td>
<td>232</td>
<td>37</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The $y$ is the number of additional cases of each type of procedure that the study hospital can be producing based on the hospital’s performance relative to its benchmark hospital. Abbreviations for the outputs (columns) and inputs (rows) are given in Table 1. The equation for the weights in the second column is Eq. (13) of Ref. (3), with the weights being the ratios of the $y$ to the $\theta$ in that equation. Without rounding, the sum of the weights equals 1.042. For the super-efficient data envelopment analysis model, as we used, the sum of weights is not constrained to equal 1.0. Although in reference (3) we arbitrarily set the negative differences equal to 0, we did not do so in this Table 3 or in Table 7. The $-49\% = 100\% \times (1/1.96 - 1)$. Where 1.96 is the effective score, rounded to 2.0 in the text of the paper.
Table 4. Sensitivity of Gaps in the Study Hospital’s Procedures to Inclusion of each Hospital Contributing to the Benchmark

<table>
<thead>
<tr>
<th>Excluded hospital</th>
<th>Additional hospital(s) added?</th>
<th>AAA</th>
<th>CABG</th>
<th>Colorectal</th>
<th>Craniotomy</th>
<th>Hip replacement</th>
<th>Hysterectomy</th>
<th>Lung resection</th>
<th>Nephrectomy</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td></td>
<td>240</td>
<td>−49</td>
<td>−18</td>
<td>119</td>
<td>42</td>
<td>−49</td>
<td>−22</td>
<td>−31</td>
</tr>
<tr>
<td>Hospital A</td>
<td>Yes</td>
<td>276</td>
<td>−53</td>
<td>−19</td>
<td>10</td>
<td>3</td>
<td>−47</td>
<td>2</td>
<td>−38</td>
</tr>
<tr>
<td>Hospital B</td>
<td>Yes</td>
<td>226</td>
<td>−50</td>
<td>−16</td>
<td>122</td>
<td>47</td>
<td>−47</td>
<td>−24</td>
<td>−32</td>
</tr>
<tr>
<td>Hospital C</td>
<td>Yes</td>
<td>233</td>
<td>−50</td>
<td>−16</td>
<td>107</td>
<td>44</td>
<td>−45</td>
<td>−20</td>
<td>−31</td>
</tr>
<tr>
<td>Hospital D</td>
<td>No</td>
<td>248</td>
<td>−51</td>
<td>−19</td>
<td>70</td>
<td>21</td>
<td>−50</td>
<td>−15</td>
<td>−37</td>
</tr>
</tbody>
</table>

Values given are percentage values. The four hospitals originally contributing to the benchmark are listed on the left (Table 3). The selection of the benchmark hospitals and their weights are a direct consequence of the selection of outputs and inputs to make the study hospital appear as efficient as possible. The gap is the ratio of the difference to the current number of cases. A negative gap infers that more cases are being done than expected from the benchmark hospital(s). A positive gap shows the potential of the study hospital to increase its workload for the specialty. When Hospital D was excluded, there were only three hospitals contributing to the benchmark (Table 3). When Hospital A, B, or C was excluded, other hospitals were added. The Table shows that the magnitude of the gaps in neurological, orthopedic, and thoracic surgery are sensitive to the inclusion of Hospital A. The Results in the case study describe how these sensitivities have no impact on management decisions. Hospital A is in a small city in Iowa. The hospital had the same technology index as the study hospital, performing all types of care except transplant. Otherwise, Hospital A had fewer beds, surgeons, county surgical charges per square mile, and region surgical charges per square mile. Nevertheless, it performed more AAA, cranietomies, and hip replacements than the study hospital.

The case study is novel, because DEA was performed without hospital discharge abstract data from the study hospital’s state. The data used were financial data from the study hospital, input and output values for the study hospital (Table 1), data from other states with accessible data (1,2), and federal websites. The remainder of our paper evaluates the validity of this process, including a direct comparison of results of DEA using state discharge data versus federal data.

Validation was performed with the same years of data that were used for our previous studies of DEA applied to surgery so that we could compare our new findings to these prior results. The Pennsylvania Health Care Cost Containment Council’s 1998 data included all patients discharged during 1998 from a non-Federal Pennsylvania hospital (1). The corresponding data for Iowa were the 2001 discharges from the Iowa Hospital Association (2,3).

Hospitals for which DEA would not be useful were excluded (1), based on two criteria.

First, hospitals were excluded that have fewer than 200 beds. Such hospitals perform too few inpatient cases. For example, in Iowa, no hospital with <200 beds performed (2) Craniotomy or Nephrectomy and only one such hospital performed CABG or Lung resection. Note that this exclusion of hospitals refers only to the hospital’s use as benchmarks. When state discharge abstract data are used to estimate the County and Region variables, all hospitals in the state are used for that purpose.

Second, hospitals were excluded that are located in counties with populations exceeding 1 million. Hospitals in large metropolitan areas (e.g., Philadelphia and Pittsburgh) have an unlimited supply of potential patients. Their challenges to growth are competition, marketing, and high fixed costs (e.g., real estate).

Figure 4 shows how DEA was performed using US federal data instead of discharge abstract data from the state of the study hospital. The federal data was obtained from public websites. The US federal data are from nonfederal short-term, general, and other specialty hospitals, just like the hospitals in the state discharge abstract databases. The steps in Figure 4 are considered below.

We developed the methodology of Figure 4 to overcome the practical limitations of the DEA discussed in the Introduction. Only the County and Region variables from Table 1 were changed. We will establish that the change does not reduce the validity of DEA (1) by showing a statistically significant, albeit modest, increase in validity of DEA results.
The Region of a hospital refers to the surgical charges for residents of the county of the hospital and of the counties contiguous to the county of the hospital (Table 1). The charges are used to provide the average relative resource use for the eight different procedures studied (Table 1) (1,2). When data are obtained from a state inpatient database, surgical procedures performed on residents who travel outside the state for care are not included. The use of federal website data overcomes this limitation.

Figure 5 shows, for each county in Iowa and Pennsylvania, the ratio of the weighted surgical charges of its region’s residents to that of just its own residents. The data were obtained just from the state databases. The ratios before normalization are all larger than 1.0, because each county’s region includes the county. The ratios were significantly less for counties contiguous to a state border (Mann–Whitney $P = 0.0002$, $n = 166$ counties). The ratios were less for border counties, because their numerators were underestimates of actual regional surgical charges. In fact, if a county’s region were homogeneous, round, and equally straddling two states, then the Region variable estimated using data from one state would be half its true value. The lower panel of Figure 5 shows that normalizing the County and Region surgical charges by land square miles successfully eliminated the bias ($P = 0.58$). The land square miles of each county comes from the US census bureau website (Fig. 4).

Figure 6 shows calculated percentage differences for county surgical charges per square mile estimated using the federal websites versus state databases. As predicted, charges were significantly larger when estimated using federal data versus state data (Wilcoxon’s signed rank test $P < 0.0001$, $n = 166$ counties). Differences were significantly larger for border counties than for non-border counties (Mann–Whitney $P = 0.001$).

Not all border counties have the same chance of its residents leaving the state for surgery. During the 2000 US Census, the place of work was determined for a sample of workers 16 yr and older residing in each county. If many residents of a county commuted
out-of-state for work, that would suggest economic activity (including healthcare) outside the state, but close to the border. Each increase in the percentage of county residents working outside the state was associated with an increase in the difference in county surgical charges per square mile estimated using the federal websites versus the state inpatient databases (Spearman’s $P < 0.0001$).

The preceding results for Figure 6 are consistent with the hypothesis that state inpatient databases can underestimate surgical rates for counties close to borders. Results should be similar, but less significant, by region. The lower panels show that our findings matched this hypothesis.

Table 5 shows that the inaccuracies in estimating the $County$ and $Region$ variables for hospitals in border counties may have influenced results of DEA for more
Table 6. Validity of the Three Methods of Data Envelopment Analysis Assessed Using Differences Between the Efficient and Inefficient Pennsylvania Hospitals

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Efficient hospitals (N = 25) (median ± quartile deviation or %)</th>
<th>Inefficient hospitals (N = 28) (median ± quartile deviation or %)</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Market share among its surgeons</td>
<td>90% ± 11%</td>
<td>64% ± 15%</td>
<td>0.003</td>
</tr>
<tr>
<td>Affiliations per surgeon</td>
<td>1.4 ± 0.2</td>
<td>1.8 ± 0.2</td>
<td>0.002</td>
</tr>
<tr>
<td>Beds</td>
<td>275 ± 109</td>
<td>263 ± 50</td>
<td>0.226</td>
</tr>
<tr>
<td>Technology</td>
<td>5 ± 2</td>
<td>5 ± 1</td>
<td>0.500</td>
</tr>
<tr>
<td>Surgeons</td>
<td>70 ± 30</td>
<td>58 ± 22</td>
<td>0.310</td>
</tr>
<tr>
<td>Border county?</td>
<td>44%</td>
<td>46%</td>
<td>0.999</td>
</tr>
<tr>
<td>Rural county?</td>
<td>24%</td>
<td>14%</td>
<td>0.488</td>
</tr>
<tr>
<td>Teaching hospital?</td>
<td>20%</td>
<td>11%</td>
<td>0.453</td>
</tr>
</tbody>
</table>

Non-significant differences support validity of the DEA, with each of the three DEA methods specified by the row

- County
  - PA data | $82 ± $47 (million) | $95 ± $41 (million) | 0.383 |
  - PA data, per square mile | $125 ± $56 (thousand) | $142 ± $97 (thousand) | 0.345 |
  - US data, per square mile | $131 ± $47 (thousand) | $157 ± $110 (thousand) | 0.254 |
- Region (i.e., county and contiguous counties)
  - PA data | $568 ± $386 (million) | $687 ± $277 (million) | 0.187 |
  - PA data, per square mile | $98 ± $70 (thousand) | $151 ± $103 (thousand) | 0.154 |
  - US data, per square mile | $103 ± $78 (thousand) | $149 ± $115 (thousand) | 0.175 |

DEA stands for "Data Envelopment Analysis". The three methods of DEA were: (i) State of Pennsylvania (PA) data without normalizing County and Region by land square miles, (ii) State of PA data with normalizing of County and Region by land square miles, and (iii) federal and PA data with normalization of County and Region by land square miles. The five inputs used in the DEA and their units are given in Table 1. The "Market share among the surgeons" had a numerator of the number of the eight studied procedures that were performed at a given hospital and a denominator of the number of such procedures performed at all hospitals by the hospital's surgeons. The "Affiliations per surgeon" was defined as the number of hospitals at which the surgeon performed at least one of the eight, studied, procedures. Exact P-values for discrete variables were obtained using Fisher's exact test.

than 90% of hospitals. More than three quarters of hospitals in Iowa and Pennsylvania with at least 200 Beds are in a region coincident to a state border. However, even that result understates the effect. A hospital that is located neither in a border county nor in a region coincident to a state border may choose at least one benchmark hospital that is located near a state border.

Although Figures 5 and 6 and Table 5 suggest benefits of using federal data for the DEA, the use of federal data should be balanced against methodological concerns. Figure 4 shows that the County and Region variables are estimated by multiplying each county’s population for 10 combinations of age and gender by the corresponding age and gender adjusted national surgical rates. The federal surgical rates will be biased estimators for the surgical rates of the residents of some counties.

The assumption of homogeneity of surgical rates among counties may have a small effect on results of the DEA. What affects the DEA is the hospital surgical workload, determined by numbers of patients, not the rates. The County and Region variables depend on the product of population and rate for each of the procedures. Among the 166 counties, the populations varied 340-fold and the population densities 1100-fold. Among the 45 counties with at least one of the hospitals studied using DEA, populations varied 41-fold and the population densities 71-fold. Variations in surgical rates were much smaller (e.g., 1.6-fold for Colectomy and 3.6-fold for Hysterectomy, when limited to benign disease for which there is high variation in practice) (22). Thus, variation in population among counties is so large that variation in surgical rates may have a negligible impact on variation in the County and Region variables among counties. We addressed this issue by reestablishing the validity of DEA using the revised methodology of Figure 4.

The six criteria applied previously to validate DEA were repeated.

1. DEA chooses for a study hospital the collection of outputs and inputs that make the study hospital appear as efficient as possible. If there is at least one other hospital that produces more of the selected outputs with the selected inputs, then the study hospital is inefficient. Each hospital is compared to all other hospitals. This process is presented graphically in Ref. (2). Efficient and inefficient hospitals should differ significantly on other measures that would indicate relatively large production of surgery (1). For example, surgeons would be more likely to practice exclusively at an efficient hospital. That may...
be because the OR management helps the surgeons get more surgery done or because each surgeon doing most of her cases at the hospital results in each surgeon filling an OR on the days she operates. Regardless of the reason why, Table 6 shows that these results hold for Pennsylvania, the state with the ancillary data (1).

2. There should not be statistically significant differences in input variables between efficient and inefficient hospitals (1). For example, efficient hospitals should not have more beds than inefficient hospitals. This is known in the DEA literature as the mathematical assumption of constant returns-to-scale. Table 6 shows that this condition holds.

3. Efficient and inefficient hospitals should match on input variables excluded from the DEA model. Previously, reasonable additional inputs for which differences were appropriately shown to be absent were binary variables: (a) whether a hospital was a teaching hospital and (b) whether a hospital was located in a rural county (1). Table 6 shows that these two conditions hold. Table 6 also shows that hospitals in border counties were, appropriately, not more or less likely to be efficient.

4. A gap is the number of additional cases of a type of procedure that the hospital can be producing based on the hospital’s performance relative to its benchmark hospital. Calculated gaps in each hospital’s cases for each type of procedure should be correlated to estimates obtained using unrelated statistical methods. A nonlinear regression model was applied to Pennsylvania data (1). Table 7 shows that this condition holds.

As a further test, calculated gaps should have face validity. There is no gold standard to which a person can mentally compare a gap when it is zero or positively valued for a hospital performing the procedure (i.e., when the gap is useful). The DEA addresses a problem for which there is no other developed methodology. On the other hand, if a hospital hardly performs a procedure, the gap value should not be negative and non-negligible. Thresholds used for a hospital “hardly performing a procedure” were fewer than 10 discharges per year of AAA, Craniotomy, Lung resection, or Nephrectomy, and fewer than 20 discharges per year of CABG, Colorectal resection, Hip replacement, and Hysterectomy (1). The nonlinear regression effectively considered the hospital’s potential elective surgical workload for each procedure to be proportional to the number of age and gender adjusted hospital admissions (1). The three methods of Data Envelopment Analysis (DEA) were: (i) State of Pennsylvania (PA) data without normalizing County and Region by land square miles, (ii) State of PA data with normalization of County and Region by land square miles, and (iii) federal and PA data with normalization of County and Region by land square miles. Rank correlations were used to compare the residuals of the nonlinear regression to the gaps of the three methods of DEA. Rank correlations were used, because there was: (i) absence of a theoretical argument for linear relationship, (ii) skewed data, and (iii) multiple zero values for gaps. The P-values were calculated using Monte-Carlo simulation to an accuracy of within 0.0001 (StatXact-7). The P-values are one-sided as only positive correlation would support the validity of the DEA.

Table 7. Validity of the Three Methods of Data Envelopment Analysis Based on Positive Spearman Rank Correlations Between Gaps in the Procedures and Estimates from Nonlinear Regressions of Hospital’s Deficits in Those Procedures

<table>
<thead>
<tr>
<th>Procedure</th>
<th>PA data</th>
<th>PA data, normalizing by land square miles</th>
<th>US data, normalizing by land square miles</th>
<th>Hospitals above threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Correlation coefficient</td>
<td>One-sided P-value</td>
<td>Correlation coefficient</td>
<td>One-sided P-value</td>
</tr>
<tr>
<td>AAA</td>
<td>0.376</td>
<td>0.010</td>
<td>0.368</td>
<td>0.010</td>
</tr>
<tr>
<td>CABG</td>
<td>0.755</td>
<td>&lt;0.001</td>
<td>0.785</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Colorectal</td>
<td>0.555</td>
<td>&lt;0.001</td>
<td>0.627</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Craniotomy</td>
<td>0.433</td>
<td>0.008</td>
<td>0.516</td>
<td>0.001</td>
</tr>
<tr>
<td>Hip replacement</td>
<td>0.632</td>
<td>&lt;0.001</td>
<td>0.609</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Hysterectomy</td>
<td>0.684</td>
<td>&lt;0.001</td>
<td>0.740</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Lung resection</td>
<td>0.563</td>
<td>&lt;0.001</td>
<td>0.421</td>
<td>0.004</td>
</tr>
<tr>
<td>Nephrectomy</td>
<td>0.484</td>
<td>0.002</td>
<td>0.412</td>
<td>0.006</td>
</tr>
</tbody>
</table>

Full procedure names and descriptions are given in Table 1. There are N = 53 Pennsylvania hospitals. Hospitals excluded for each procedure were those “hardly performing a procedure,” defined as fewer than 10 discharges per year of AAA, Craniotomy, Lung resection, or Nephrectomy, and fewer than 20 discharges per year of CABG, Colorectal resection, Hip replacement, and Hysterectomy (1). The nonlinear regression effectively considered the hospital’s potential elective surgical workload for each procedure to be proportional to the number of age and gender adjusted hospital admissions (1). The three methods of Data Envelopment Analysis (DEA) were: (i) State of Pennsylvania (PA) data without normalizing County and Region by land square miles, (ii) State of PA data with normalization of County and Region by land square miles, and (iii) federal and PA data with normalization of County and Region by land square miles. Rank correlations were used to compare the residuals of the nonlinear regression to the gaps of the three methods of DEA. Rank correlations were used, because there was: (i) absence of a theoretical argument for linear relationship, (ii) skewed data, and (iii) multiple zero values for gaps. The P-values were calculated using Monte-Carlo simulation to an accuracy of within 0.0001 (StatXact-7). The P-values are one-sided as only positive correlation would support the validity of the DEA.
Among states (2). The county sizes had been different. Making small changes to the methodology without changing the sample size should not affect the efficiency scores between different DEA models. Changes in efficiency scores between different DEA models can be evaluated by considering pair-wise differences in efficiency scores (23). For the few hospitals with large decreases in efficiency scores, qualitative assessments should indicate that the changes likely represent improvements in accuracy. It was now being compared to a broader population of potential benchmark hospitals (23). The bottom half of Table 8 shows that this prediction holds.

Previously we tried unsuccessfully to pool data among states (2). The county sizes had been different. Now, County and Region are normalized by land square miles. Federal data are used to compensate for border effects. Hospitals are limited to those with enough Beds for DEA to be relevant, eliminating multiple small hospitals in Iowa.

Pooling hospitals among states should result in small but statistically significant reductions in the median efficiency scores, because each hospital is now being compared to a broader population of potential benchmark hospitals (23). The bottom half of Table 8 shows that this prediction holds.

Increasing the sample size can only result in declines in efficiency scores (23). For the few hospitals with large decreases in efficiency scores, qualitative assessments should indicate that the changes likely represent improvements in accuracy. Strikingly, the two hospitals with the largest declines were the hospitals with the two highest original scores: 7.75–1.37 and 6.96–4.22. This observation further supports the validity of pooling hospitals between states. As the number of hospitals increases, the likelihood of encountering another similar hospital that is close to being as efficient as possible at producing surgery also increases (23). The two biggest outliers were reigned in by the changes to the methodology. The incremental effect of adding an additional hospital tends to decrease as the sample size increases (23). None of the other declines in efficiency scores exceeded 0.6.

The hospital with the largest decline was referred to as Example Hospital C in Ref. (1). Its unique feature was that although its county had the smallest population and amount of surgery among its residents, the hospital’s Craniotomy caseload was the third most and CABG caseload was the seventh most among hospitals in Pennsylvania. We used this hospital to highlight that market visibility (Beds, Surgeons, and Technology) need not be an important cause of limitations on the capture of the surgical market (1). In the current paper, the decline in the hospital’s efficiency score occurred as a result of normalization by land square miles. The hospital’s county not only had the smallest population and weighted surgical charges, but was also the county with the smallest land square miles. Even though the analysis continues to show the hospital is unique, the original results were likely inflated. This finding supports the face validity of the modification of the DEA methodology shown in Figure 4.

The hospital with the second largest decline in its efficiency score was the study hospital in Refs. (2) and (3). As described previously, the input chosen to make the hospital appear as efficient as possible was its County weighted surgical charges, just as for the preceding hospital. The productivity ratios presented in Ref. (3) showed that although the hospital’s gaps equaled zero for all outputs in the original model, the hospital was relatively weak in orthopedics (Hip replacement). The second weakest specialty was cardiac surgery (CABG). We used that finding to highlight the usefulness of examining the productivity ratios (3). In the current paper, the decline in the hospital’s efficiency score resulted from the consolidation of data between states. The reason was that this Iowa hospital was now being compared to the Pennsylvania hospital in the preceding paragraph, with its strength in CABG. The result was a slight gap for the Iowa hospital in CABG (23 cases) and Hip replacement (25 cases). These results, too, seem reasonable.

### Table 8. Pair-wise Differences in Efficiency Scores for Hospitals Among Different Data Envelopment Analysis Models

<table>
<thead>
<tr>
<th>Modification of methodology</th>
<th>State</th>
<th>Original median efficiency score</th>
<th>Median change</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonsignificant median changes in efficiency scores support validity of the modification of methodology</td>
<td>Iowa</td>
<td>1.265 (N = 21)</td>
<td>−0.007</td>
<td>0.855</td>
</tr>
<tr>
<td>Normalizing and region by land and square miles</td>
<td>Pennsylvania</td>
<td>0.985 (N = 53)</td>
<td>0.007</td>
<td>0.627</td>
</tr>
<tr>
<td>Normalizing and using US data</td>
<td>Iowa</td>
<td>1.102 (N = 74)</td>
<td>0.006</td>
<td>0.647</td>
</tr>
<tr>
<td>Normalizing and region by land and square miles</td>
<td>Iowa</td>
<td>1.102 (N = 74)</td>
<td>0.005</td>
<td>0.559</td>
</tr>
<tr>
<td>Pooling data from Iowa and Pennsylvania</td>
<td>Both</td>
<td></td>
<td>0.005</td>
<td>0.559</td>
</tr>
<tr>
<td>Significant median changes of predicted magnitude support validity of the modification of methodology</td>
<td>Iowa</td>
<td>−0.094</td>
<td>0.026</td>
<td></td>
</tr>
<tr>
<td>Pooling data from Iowa and Pennsylvania</td>
<td>Both</td>
<td>−0.061</td>
<td>&lt;0.001</td>
<td></td>
</tr>
<tr>
<td>Pooling data from Iowa and Pennsylvania</td>
<td>Both</td>
<td>−0.067</td>
<td>&lt;0.001</td>
<td></td>
</tr>
</tbody>
</table>

The Wilcoxon’s signed rank test was used. P-values are exact (StatFact-7).
DISCUSSION

We applied and validated new methods for using DEA to judge the ability of a hospital to perform additional inpatient surgery for specific specialties. We then applied the DEA to a hospital in a different state to assess decisions related to tactical selective increases in OR block time. The results were presented in reverse order to emphasize the practicality of the method for a real hospital and to avoid a separate background section with hypothetical examples.

The ability to use federal website data to assess County and Region surgical workload makes DEA results broadly practical and accessible for OR managers. Each hospital wanting to assess its inpatient surgical workload can compare itself to other hospitals that have previously been examined. As each hospital benefits, it contributes its own data, making DEA that much more informative to the next hospital, and so forth.

Although the use of federal website data increases the validity of DEA by a statistically significant amount, the increase is modest. We believe that the increase in validity is of importance only to the extent that it is very strong evidence that the validity was not diminished (i.e., the benefit of the new method is that it is easier to perform).

Our use of DEA is unique in evaluating hospital performance (24). Almost all previous studies of hospital efficiency using DEA have been analyses of data from multiple hospitals to assess health systems and policy (24). A few studies have applied DEA to data from individual organizations to support managerial decision-making within the organization (e.g., physicians within a large medical group). In contrast, our focus has been the application of DEA to data from multiple hospitals to support managerial decision-making within a single hospital (1–3).

Our case study addressed one application of DEA: long-term (1 yr) tactical allocation of additional OR block time (10). We previously used the hospital in the last paragraph of the Validation section to show other similar applications. That hospital performed 40% of the neurosurgery and 25% of the inpatient urologic surgery in Iowa (2). Discharges in those specialties were twice that of the next highest hospitals. In contrast, the hospital performed 9% of Iowa’s cardiac surgery. Its cardiac workload was half that of the highest volume hospital. One question asked at the hospital was whether cardiac surgery could be increased by the overlapping finishing of a case in one OR with the induction of the next patient in another OR to reduce turnover times. Another question asked was whether capital investment for minimally invasive equipment could prompt additional referral of patients resulting in growth in cardiac surgery. The current paper shows that may be caseload could have been increased by a negligible two CABG patients per month. The limiting factor was, and still is, a lack of elderly patients living near the hospital. This example shows how DEA can be of value to anesthesia groups in recruitment decisions. DEA can also be useful for OR managers and surgical services committees with discretion in how increases in operational budgets are allocated and how capital purchases are chosen.

The case study shows that we are making progress at understanding the steps and issues in improving decision-making for allocating additional OR time tactically. Figures 1–4 show that decision-makers are very unlikely to know whether a haphazard alternative method of analysis is giving an answer that is “close enough” without also doing formal, coupled financial and market growth (DEA) analyses. Our interpretation is strengthened by the fact that simple versions of these analyses were sufficient for our study hospital. Administrators at the hospital happened not to need to consider limited ICU beds (16), uncertainty in estimated contribution margins per OR hour (6,9), expert estimates of market growth potential (10), or the competitive effects of other hospitals (2,25). Having learned what it takes to model surgical demand appropriately (Fig. 4), we are skeptical that intuition alone can come close to getting the correct answer.

An appropriate alternative to performing a mathematical analysis to assist in the tactical decision is not to make a tactical decision. Instead of reserving additional OR time for a few targeted specialties, the extra OR time would be used as overflow. The overflow OR time is used by specialties that have filled their allocated OR time (12,13) and have more cases to perform (14,15). Over many months, the service that consistently uses the overflow OR time would have some of the overflow OR time allocated to them (13,15). Essentially, the allocation of OR time to specific services is being limited to the shorter-term operational stage. An advantage to this “wait and see” approach is that experienced managers generally prefer to postpone making decisions when having to choose among alternatives that require trade-offs (26). Some readers may want to refer to the case studies of the current paper and of a prior paper (10) as evidence that deferring tactical decisions can be a good choice.

In the case study (Fig. 1), consideration of incremental revenue was limited to hospital reimbursement. However, also including professional reimbursement in the analysis can provide insight into differences in perspective between hospital and physicians (27). In addition, or alternatively, an intangible value can be added for each additional patient of desired types (e.g., cancer patient receiving care at a cancer hospital or cardiac patient at a hospital making the strategic decision to focus on cardiac care). The incremental revenue may even be zero for all patients other than those targeted by a public ministry. However, the latter is increasingly uncommon as more countries migrate to prospective payment based on Diagnosis Related Groups (28).

The hospital studied in our paper was private. Nonetheless, the linkage of financial analysis with DEA for purposes of tactical increases in allocated OR
time can be useful for public hospitals. Unless every purchase request by every surgeon, and every request for salary support for new surgeons is satisfied, the hospital makes tactical decisions based in part on financial criteria. The allocation of additional OR time tactically is a budgetary decision, just like the purchase of new expensive OR equipment to care for a new population of patients.

A finding of the study hospital was unusual. Specialties with some of the highest hospital contribution margins per OR hour were those doing outpatient surgery (e.g., pain medicine) (Fig. 2). The finding reinforces that contribution margin per OR hour results by specialty differ markedly among facilities (27). Usually, the highest values are for specialties with substantial inpatient surgery (4,16,29). Specialties with longer average lengths of stays can have higher contribution margins per OR hour, because such patients offer more opportunities for discharge of patients a few days sooner than average (27). To explain financial analysis results for individual surgeons (specialties), we use the difference between each inpatient’s hospital length of stay and the national average length of stay for all patients with the same Diagnosis Related Groups (27,30).

For the studied hospital, data on contribution margin per OR hour (Figs. 1 and 2) did not need to be combined mathematically with estimates of potential market growth. If they had, we would have used equation (19) in Ref. (10). A question that administrators and surgeons typically ask is what happens if more OR time were allocated tactically to a specialty and they do not use it. The OR time that would otherwise be under-utilized would be released for services that have filled their allocated OR time and still have more cases to schedule (13,14,31). The important implication is that there is marked mitigation of the financial risk of inaccurate results from DEA (10).

Although the method of performing DEA using Figure 4 enables rapid decisions for some problems, the approach seems limited for determining the cause of a gap value. For example, Tables 2–4 show that the hospital in the case study had a weakness in vascular surgery. This was no surprise, considering that the hospital had only one vascular surgeon. What was unknown was the potential financial return from performing more vascular surgery. Nevertheless, if the hospital had recruited another vascular surgeon, would the gap have been reduced or would two surgeons be doing the same relatively small numbers of cases? Those questions can be addressed by considering the impact of competing hospitals on the study hospital (2,25) and by studying the types of vascular procedures performed at the study hospital and their differentiation from that of competing hospitals (32,33). Both assessments rely on the state data that we bypassed by following the methodology outlined in Figure 4. Thus, state discharge data provide information useful to individual hospitals that cannot be completely replaced by federal data. Such uses of the state data are not affected by the issues of the border counties considered in Figures 5 and 6.

Finally, DEA is limited to the assessment of workload at hospitals, which may not be relevant to medical practices. If an anesthesia group practices at two nearby hospitals, both with gaps in a specialty, the actions taken by the hospitals to address the gaps may be irrelevant to the group, since the numbers of patients at the two hospitals combined could remain unchanged.

ACKNOWLEDGMENTS
The authors thank Ruth Wachtel, PhD MBA, for her thoughtful review of our paper.

REFERENCES
13. Dexter F, Traub RD, Macario A. How to release allocated operating room time to increase efficiency. Predicting which surgeons will have the most under-utilized operating room time. Anesth Analg 2003;96:507–12.

ERRATUM

In the February 2007 issue of Anesthesia & Analgesia, in the article by O’Neill and Dexter, “Tactical Increases in Operating Room Block Time Based on Financial Data and Market Growth Estimates from Data Envelopment Analysis” (Anesth Analg 2007;104:355–68), there are two errors in the footnote to Table 3 on page 359. The first sentence of the caption starts: “The % is the number of additional cases” which should be “The gap is the number of additional cases.” The second line has “with the weights being the ratios of the v to θ, which should be “with the weights being the ratios of the λ to θ.” The publisher regrets the error.