HARVARD Kennedy School
JOHN F. KENNEDY SCHOOL OF GOVERNMENT

Who is an Efficient and Effective Physician? Evidence from Emergence Medicine

Faculty Research Working Paper Series

Soroush Saghafian
Harvard Kennedy School

Raha Imanirad
Harvard Business School

Stephen J. Traub
Mayo Clinic Arizona

September 2018

RWP18-029

Who is an Efficient and Effective Physician? Evidence from Emergence Medicine

Soroush Saghafian,[1] Raha Imanirad,[2] Stephen J. Traub[3]

[1]Harvard Kennedy School, Harvard University, Cambridge, MA

[2]Technology and Operations Management, Harvard Business School, Cambridge, MA

[3]Department of Emergency Medicine, Mayo Clinic Arizona, Phoenix, AZ

Abstract

Improving the performance of the healthcare sector requires an understanding of the efficiency and effectiveness of care delivered by providers. Although this topic is of great interest to policymakers, researchers, and hospital managers, fair and scientific methods of measuring efficiency and effectiveness of care delivery have proven elusive. Through Data Envelopment Analysis (DEA), we make use of evidence from care delivered by emergency physicians, and shed light on scientific metrics that can gauge performance in terms of efficiency and effectiveness. We use these metrics along with Machine Learning techniques and Tobit analyses to identify the distinguishing behaviors of physicians who perform highly on these metrics. Our findings indicate a statistically significant positive relationship between a physician's effectiveness and efficiency scores suggesting that, contrary to conventional wisdom, high levels of effectiveness are not necessarily associated with low efficiency levels. In addition, we find that a physician's effectiveness is positively associated with his/her average contact-to-disposition time and negatively associated with his/her years of experience. We also find a statistically significant negative relationship between a physician's efficiency and his/her average MRI orders per patient visit. Furthermore, we find evidence of a peer effect of one

physician upon another, which suggests an opportunity to improve system performance by taking physician characteristics into account when determining the set of physicians that should be scheduled during same shifts.

Introduction

Healthcare spending is projected to rise to 19.9% of the GDP by 2025 (Keehan et al. 2017), spurring interest in finding new ways to increase both the efficiency and effectiveness of care delivery. As most decisions regarding utilization of healthcare services are ultimately made by frontline clinicians (Tsugawa et al. 2017), understanding and evaluating provider performance in a fair and scientific manner could help identify sources of waste in healthcare spending and help to determine optimal incentives for healthcare reimbursement. Although care delivery performance measurement initiatives have proliferated in recent years, there are few rigorous scientific methods to evaluate the efficiency and effectiveness of physicians.

We employ Data Envelopment Analysis (DEA)—a linear programming (LP) optimization technique that provides a multi-dimensional evaluation tool—to develop and evaluate metrics related to both efficiency and effectiveness. Efficiency in the Emergency Department (ED) can be measured in multiple ways, but throughput—the average number of patients seen by a provider per unit of time—per resources used by the provider possesses significant face validity as an output-to-input ratio for exploratory analysis. For a given level of resources used by the provider, a higher throughput means that more patients can be moved through the ED per unit of time. Given that ED crowding has reached epidemic proportions in the last several years (Salway et al. 2017), improving throughput without adding resources has become even more important.

Similarly, effectiveness of care delivery in the ED can be measured in different ways. However, compared to efficiency, it is a more inherently difficult outcome to measure, and hence, we suggest using a composite of three metrics. The first pertains only to discharged patients. Specifically, for a fixed level of resources used by a provider, we consider the percentage of discharged patients who do not return to the ED within 72 hours. Returns to the ED within 72 hours of discharge may result from a sub-optimal (i.e., ineffective) first visit, in which not all medical issues were sufficiently identified or addressed. The 72-hour rate of return has been even proposed as a measure of quality in the Emergency Medicine literature, although using it for measuring quality is controversial (Abualenain et al. 2013, Pham et al. 2011, Klasco et al. 2015). The second and third metrics pertain only to admitted patients (i.e., those admitted to the hospital after their ED visit). Specifically, we consider how frequently a physician admits patients who are subsequently discharged after a brief period, which suggests that the physician potentially *overcalled* the patients' illness severity (i.e., patients are admitted when they could have potentially been discharged). We also consider how frequently a physician admits patients whose care is escalated from a low-acuity bed (ward/floor) to a high-acuity bed (Intermediate Unit or Intensive Care Unit), which suggests that the physician potentially *undercalled* the patients' illness severity. Although none of these three metrics is ideal in isolation in measuring effectiveness of care delivery, their use in the aggregate (per resources used by the provider) possesses face validity, and covers the performance of a physician among both discharged and admitted patients.

We apply our proposed DEA approach (which considers the above measures as outputs and the level of resources used by the physician as input) to a large data set of care delivered by ED physicians that includes more than 115,000 patient visits. To learn about what the high-performing physicians do differently from other physicians, and thereby generate insights into best practices, we first make use of some Machine Learning (ML) algorithms (including k-means, spectral, and random forest) to (a) segment physicians based on their effectiveness and efficiency scores, and (b) identify physicians who have a high performance. We then conduct a second-stage analysis in which we use a Tobit framework to identify factors (e.g., test order count, physician contact-to-disposition time, etc.) associated with higher levels of performance. We also use our framework to study peer effect: the effect of the characteristics of another physician who is scheduled to work side-by-side with the first physician. In particular, we consider peer physician characteristics such as relative effectiveness, efficiency, experience, gender, and type of medical degree (MD vs. DO), and examine how they affect another physician's efficiency and effectiveness.

Our results offer important insights into physician performance. Contrary to the conventional wisdom that efficiency may come at the price of effectiveness (and vice-versa), our findings demonstrate a statistically significant positive association between efficiency and effectiveness (P = 0.0209). We also find that a physician's efficiency is negatively associated with his/her average number of MRI orders per patient visit (P = 0.0496). In addition, we observe a statistically significant negative relationship between a physician's effectiveness and his/her experience level measured in years since graduation from medical school (P = 0.0301), which can be due to the recent changes in medical training programs that further emphasize the efficient use of hospital resources. We also identify a statistically significant positive relationship between a physician's effectiveness and his/her average contact-to-disposition time (P = 0.0290), suggesting that effectiveness can be enhanced via training programs that enable physicians to avoid speed-up activities that reduce the contact-to-disposition time but negatively affect the effectiveness of the care delivered (e.g., cutting necessary tests or reducing value added direct or indirect care time).

Finally, our findings with regards to physician peer effects suggest a statistically significant impact of peer effect on each individual physician's performance, a fact that can be utilized by hospital administrators to improve performance via scheduling the correct set of physicians during same shifts. In particular, our findings suggest that working alongside a more effective peer is positively associated with improving both a physician's efficiency (P = 0.0044) and effectiveness (P = 0.0147).

Data

Our data consist of detailed care delivery information in a leading U.S. hospital with 32 ED physicians. The patients are algorithmically assigned to physicians upon arrival through an automated rotational patient assignment process (Traub et al. 2016). This workflow essentially removes all patient selection biases (preferences of physicians who may "cherry-pick" their patients), providing a unique opportunity to treat performance outcomes as almost entirely natural experiments. We included all patient visits from July 12, 2012, to July 31, 2016 who were identified in the electronic health record system as having been seen by an ED physician. Patient specific-data included demographic (age, gender, race, etc.) and insurance information, among others. Encounter-level data included laboratory tests, chief complaint, Emergency Severity Index (ESI), day of the ED visit, time of day, etc., totaling over 70 variables. A complete list of these variables is available in Electronic Companion EC.1. To avoid distortion of the results by outliers, 6 physicians with relatively low patient volumes (fewer than 250 visits over the 4-year period) were excluded from the analysis.

DEA Models

DEA, first introduced by Charnes, Cooper, and Rhodes in (Charnes et al. 1978), is a methodology useful in evaluating the relative performance of a set of decision making units (DMUs) in a multiple input, multiple output setting. A DMU can be viewed as an entity responsible for converting a number of inputs into a set of outputs and whose performance is to be evaluated relative to its peers (Cooper et al. 2007). Contrary to a central tendency approach which evaluates units relative to an average performer, DEA computes the relative performance by using the best performing units as the basis for comparison. One of the key advantages of DEA over other regression-based statistical methods is that it does not require specification of any functional relationship (e.g., a specific linear or non-linear model) between inputs and outputs. As a result, DEA can uncover information that remains hidden from other methodologies, and hence, might capture a more complete picture of a DMU's performance relative to the resources it uses. As a data-oriented approach, however, DEA is vulnerable to data errors and outliers.

The conventional input-oriented DEA methodology evaluates each DMU j in the population based upon a set of inputs $\{x_{ij}\}_{i=1}^{I}$ and outputs $\{y_{rj}\}_{r=1}^{R}$ while assuming a proportional reduction in all inputs. In an output-oriented setting, this methodology provides for a proportional expansion in outputs rather than a reduction in inputs. For the goals of this study, we have chosen the input-oriented mechanism.

The original DEA model was based on a constant returns to scale (CRS) methodology. Banker et al. 1984 extended the CRS model to allow for variable returns to scale (VRS). The input-oriented CRS model is based on the Linear Program (LP):

$$
\begin{aligned}
max \quad & \theta = \frac{\sum_r u_r y_{rj_o}}{\sum_i \nu_i x_{ij_o}} && && (1)\\
s.t. \quad & \frac{\sum_r u_r y_{rj}}{\sum_i \nu_i x_{ij}} \leq 1 && j = 1, ..., n, \\
& u_r, \nu_i \geq 0, && r = 1, ..., R; \quad i = 1, ..., I,
\end{aligned}
$$

where y_{rj_o} and x_{ij_o} represent the output(s) and input(s) of DMU j_o, respectively, and $\{u_r\}_{r=1}^{R}$ and $\{\nu_i\}_{i=1}^{I}$ are decision variables. In order to develop scientific scores to evaluate the performance of physicians, we develop two DEA models: (1) an *effectiveness DEA model*, and (2) an *efficiency DEA model*. The effectiveness model considers the relative use of hospital resources by a physician to the percentage of patients discharged by him/her who do not return to the ED within 72 hours as well as to the percentage of patients admitted by him/her via a correct call (i.e., without an overcall or an undercall). The efficiency DEA model considers the relative use of hospital resources by a physician to his/her throughput. In these models, DMUs comprise individual physicians who use hospital resources to deliver care. The choice of the input and output variables in each model is based on the view of the physician as a "production entity" which utilizes a number of hospital resources (inputs) to generate efficient and effective care (outputs). It is important to note that in DEA there is no objective definition of the right variables to use as inputs and outputs. We have chosen herein to define the models' input and output variables in terms of parameters that (a) best reflect a physician's performance, and (b) for which there is at least face validity and some level of agreement among physicians. As robustness checks, we have repeated our analyses with different combinations of input/output variables, and have observed that the majority of our main results presented in the Discussion and Results (Section 6) hold. In addition, we used Wagner and Shimshak's (Wagner et al. 2003) stepwise variable selection algorithm to validate our choice of model variables. The results of applying this algorithm to both the effectiveness and efficiency models are provided in Electronic Companion EC.2. Finally, we note that due to the nature of the automated rotational patient assignment algorithm implemented in our partner hospital which randomly assigns arriving patients to physicians, risk-adjustments of outcome measures are likely not essential. Nevertheless, in our statistical framework, we control for various patient characteristics that might affect physician performance.

In what follows, we first describe the input and output variables for our effectiveness DEA model (Section 3.1), and then describe these variables for our efficiency DEA model (Section 3.2). These variables are then used along with the optimization program (1) to create efficiency and effectiveness scores. We also describe how we use these scores to generate insights into the role of peer effect on the performance of each individual physician (Section 3.3). Finally, we discuss our results using some ML techniques to identify highly effective and efficient physicians (Section 4) as well as our statistical methodology for learning about what such physicians do differently than other physicians (Section 5).

Effectiveness DEA Model

Outputs:

- *Rate of discharged patients who do not return within 72-hours*: Since a high 72-hour return rate is an undesirable indicator of care delivery effectiveness, we use the proportion of patients discharged by a physician who did not return to the ED within 72 hours of their original discharge as the physician's output variable.

- *Rate of admitted patients who are not discharged within 17-hours*: We choose the percentage of patients admitted by a physician to the hospital who were subsequently discharged from the hospital within 17 hours of their admission as a proxy for how often the physician potentially overcalls his/her patients' illness severity. Since this could be considered an undesirable output, we use the 17-hour non-discharge patient admission rate as an output variable.

- *Rate of patients admitted to a floor/ward bed who are not upgraded within 6-hours*: We choose the percentage of patients admitted by the physician to the hospital who were upgraded from a floor bed to an intermediate care or ICU bed within 6 hours of their admission as a proxy for how often the physician potentially undercalls his/her patients' illness severity. Since this would be considered an undesirable output, we use the 6-hour non-upgrade patient admission rate as an output variable.

  It should be noted that the first output variable above is suitable for considering performance among discharged patients, while the two other output variables consider performance among admitted patients. The choice of threshold numbers (72, 17, and 6) is made based on observations made in the literature (see, e.g., Keith et al. 1989, Gordon et al. 1998, and the references therein) and conversations with ED physicians. In addition, we perform sensitivity analyses on these thresholds by changing each of them within a range, and observe that our main results still hold.

Inputs:

- *Lab Order Count*: Average number of the physician's lab orders per patient visit;

- *Plain Radiograph Order Count*: Average number of the physician's plain radiograph orders per patient visit.

We refer to θ scores (see Eq. (1)) generated by the DEA model with the above input-output parameters as physicians' *effectiveness scores*.

Efficiency DEA Model

Output:

- *Throughput*: Average number of patients seen by the physician per shift;

Inputs:

- *Lab Order Count*: Average number of the physician's lab orders per patient visit;
- *Ultrasound Order Count*: Average number of the physician's ultrasound orders per patient visit.

It is important to note that we compared the physicians' average number of hours worked per shift using the paired-observation t-test method and after removing two physicians from our analysis, we found no significant differences between the remaining physicians' average hours worked per shift. We did not include the plain radiograph order count as an input variable in the efficiency model because of its negative correlation with the model's output variable. We considered but excluded other resource utilization inputs such as patient admission rate and other test order counts such as MRI and CT Scan because of their negative correlation with the output variable. Our choice of variables for both models was validated using the stepwise variable selection method mentioned earlier. We refer to θ scores (see Eq. (1)) generated by the DEA model with the above input-output parameters as physicians' *efficiency scores*.

Peer-Effect DEA Model

In order to also examine the effects of peer presence on a physician's effectiveness and efficiency, we use a variation of the proposed DEA models in which each DMU, denoted by

j_k, comprise a physician j who has worked alongside his/her peer physician k for at least 5 hours. We choose the 5-hour criterion to be able to capture any meaningful peer physician influence. To confirm that our results were not sensitive to the choice of the 5-hour duration, we repeated our analysis using a 3-hour criterion. We observed no significant difference in outcomes. Thus, we utilize a total of four DEA models: effectiveness and efficiency for both individual and peer effect performance.

The models presented in Sections 3.1, 3.2, and 3.3 were all tested for isotonicity. Since there is no reason to believe that an increase in inputs results in a proportional change in the outputs, a variable returns to scale model was used. This assumption was tested by Simar and Wilson's (Simar et al. 2002, Simar et al. 2011) returns-to-scale tests for input-oriented DEA models. We further assume that the quantities of inputs are treated as endogenous, and hence, are influenced by physicians. For this reason, an input-oriented approach was used to test whether a DMU under evaluation can reduce its inputs while keeping the outputs at their current levels.

Physician Clustering

Prior to examining what highly efficient and effective physicians do differently than their peers, we must first identify such physicians. That is, we need a clustering scheme to categorize physicians based on their generated effectiveness and efficiency scores (see Figure 1).

To this end, we utilized a number of clustering algorithms from the ML literature including k-means, spectral, and random forest with the goal of segmenting physicians into the following four groups based on their scores:

Group 1. Higher effectiveness / Higher efficiency

Group 2. Higher efficiency / Lower effectiveness

Group 3. Higher effectiveness / Lower efficiency

Group 4. Lower efficiency/ Lower effectiveness

Figure 2 demonstrates how physicians are partitioned into the aforementioned clusters using the k-means algorithm, which creates groups based on the euclidean distance between the data points. Figure 3 presents the clustering of physicians based on the random forest algorithm, and Figure 4 illustrates the results based on the spectral clustering algorithm.

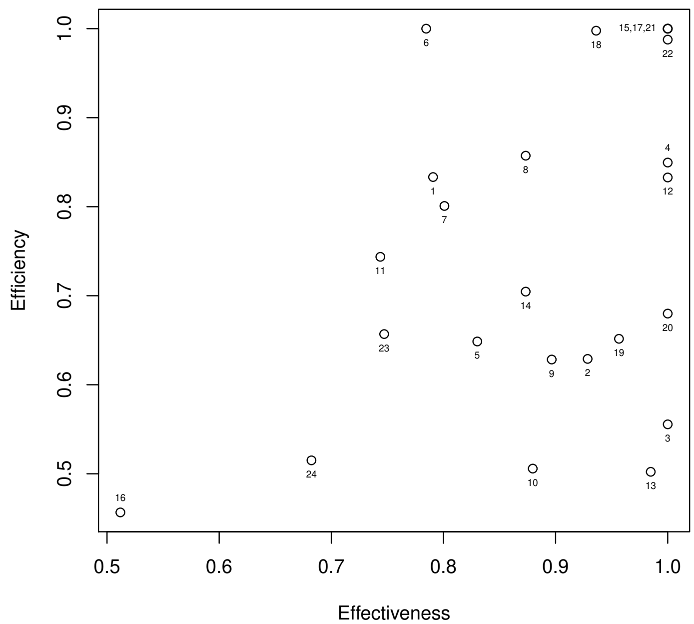

Figure 1: Physicians' Performance Scores

We compared the performance of all three algorithms using the Silhouette clustering evaluation criterion (Rousseuw 1987), which provides an index measuring both cluster cohesion and separation, and the multi-class Area Under the Curve (AUC) measure, as defined by Hand et al. 2001. Using these two criteria, we concluded that the k-means algorithm provides the most accurate clustering. Thus, we follow the partitioning of physicians into the four above-mentioned groups based on the result of the k-means algorithm (Figure 2).

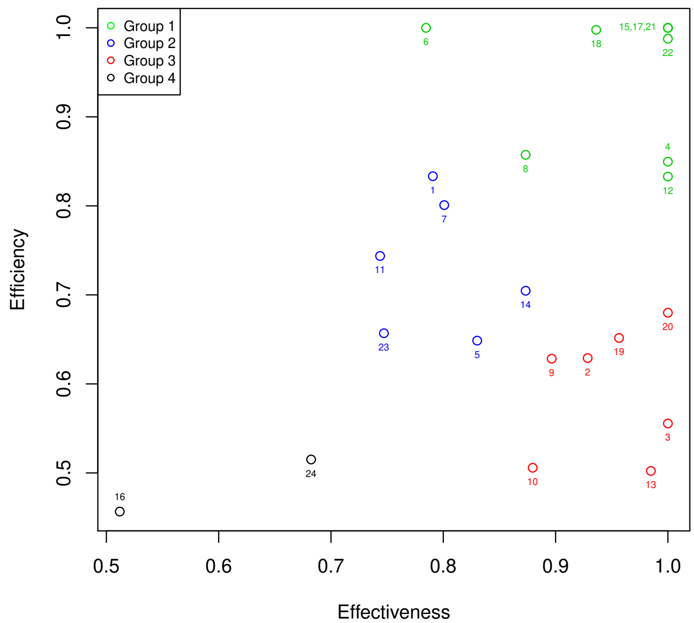

Figure 2: K-means Clustering - Silhouette coefficient = 0.6890, AUC = 0.8533

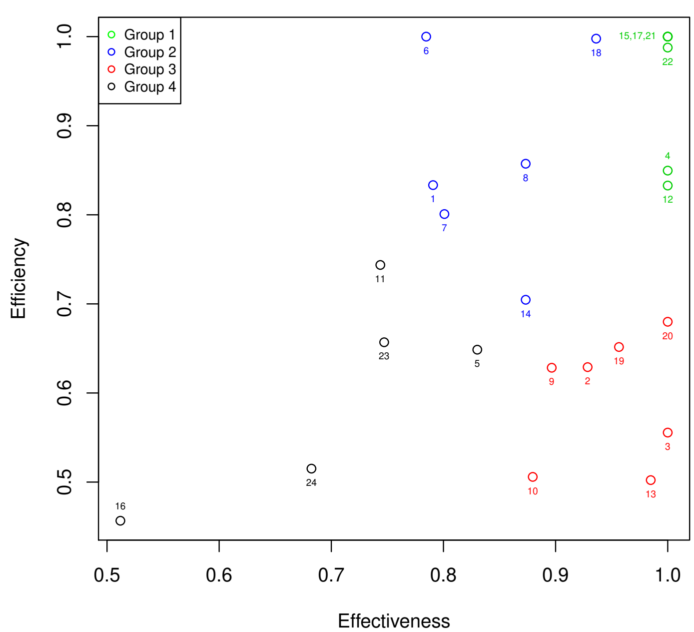

Figure 3: Random Forest Clustering - Silhouette coefficient = 0.6790, AUC = 0.8487

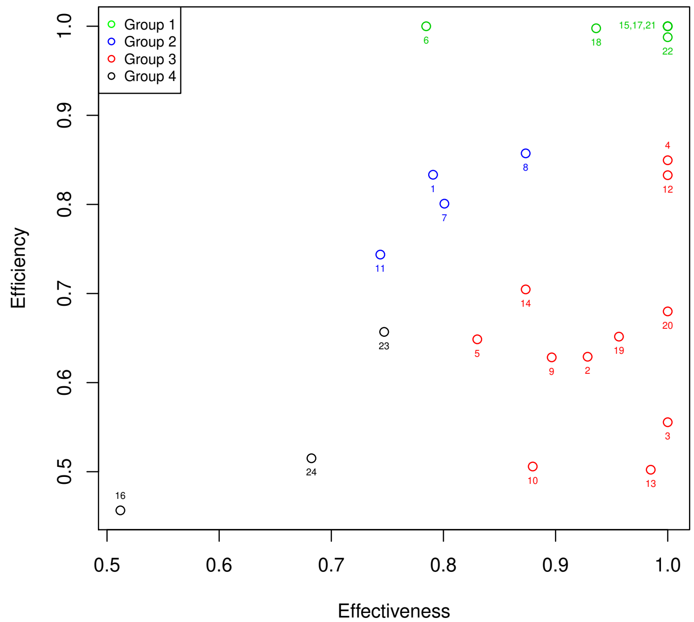

Figure 4: Spectral Clustering - Silhouette coefficient = 0.6650, AUC = 0.8233

Statistical Methodology

With the goal of learning about the practice of physicians who have a better performance than the rest, we regress the generated DEA scores of physician i (θ_i), on a set of explanatory variables related to patient, physician, and peer physician j characteristics which we denote by U_i, W_i, and Z_{ij}, respectively. The regression model takes the following general form

$$\theta_i = a + U_i\gamma + W_i\alpha + Z_{ij}\beta + \epsilon_i, \tag{2}$$

where ϵ_i is a statistical noise and a is a fixed effect (constant). In order to estimate the coefficients in equation (2), a regression technique other than the standard multivariate regression is needed. This is because this standard regression technique assumes a normal and homoscedastic distribution of the noise. However, since the DEA scores are between 0 and 1, our dependent variable is bounded and error terms may not satisfy these assumptions.

Tobit regression can be used whenever there is truncation, causing a mass of observations at a threshold value such as 0 or 1 (Chilingerian 1995). Although unlike the case of truncation, DEA does not exclude observations greater than 1 (or below 0), it does not allow a DMU to be assigned a value outside the range [0, 1]. Hence, DEA easily fits the requirement of the Tobit model (Chilingerian 1995). Following the normalization approach of Greene (1993) which assumes a censoring point at zero, we transform the DEA scores to:

$$y_i = (1/\theta_i) - 1,$$

where θ_i is the DEA measure of physician i's performance. The transformed DEA scores then become the dependent variable that takes the form:

$$y_i = \begin{cases} B'x_i + u_i, & \text{if } y_i > 0, \\ 0, & \text{otherwise,} \end{cases}$$

where B is a $k \times 1$ vector of coefficients and x_i is a $k \times 1$ vector of covariates, and u_i is the error term that is normally distributed with a mean of zero and a variance of σ^2. To account for unobserved serial correlation in the DEA scores, we used Simar and Wilson's bootstrap procedure (Simar et al. 1998) for bias-correction of the scores.

Discussion and Results

We begin our analysis by examining the relationship between physicians' scores on measures related to effectiveness and efficiency. Importantly, we find that higher scores on the effectiveness metric do not lead to lower scores on the efficiency metric, as conventional wisdom might suggest. Rather, there is a statistically significant positive relationship between the two scores (see Table 1): effective physicians are more likely to be efficient as well. This is an important observation, especially in the view of traditional debates that argue healthcare providers cannot become both more effective and more efficient at the same time.

As shown in Table 1, we find a statistically significant negative relationship between a physician's effectiveness score and his/her experience level. This is a counterintuitive result, which questions the traditional belief that experience enables physicians to use hospital resources more frugally. This may be due to the recent changes in medical training that require trainees to think more about efficient use of healthcare resources, though our data is insufficient in this regard.

Our results also indicate a statistically significant positive correlation between a physician's effectiveness score and his/her average contact-to-disposition time (the time difference between the first contact with the patient and issuing a disposition order for him/her), indicating that more effective physicians avoid speed-up activities (on direct or indirect care) that negatively affect the effectiveness of the care delivered. However, because of the fractured nature of cognitive processing of ED patients (physicians do not think continuously about each patient), it is not clear if this finding indicates that effective physicians spend more time thinking about each patient, see patients earlier in their ED encounter, or measure some other phenomenon altogether.

The regression results regarding individual physician efficiency are displayed in Table 2. The results indicate a statistically significant negative relationship between a physician's efficiency score and his/her average MRI orders per patient visit. We also observe a negative effect, although not statistically significant, of the average Intravenous (IV) medications and fluids administered on a physician's efficiency, suggesting that efficient physicians are those who order less tests, or (alternatively) order tests more intelligently. The fact that physicians with a lower number of ordered tests have higher scores on the efficiency metric supports a theory not only that there are inherent differences among physicians with respect to efficiency, but that the efficiency of providers might be improved via training programs that enable providers to decrease their use of unnecessary tests.

Table 1: Regression Results - Effectiveness Model - Individual Physician

Dependent variable: DEA Effectiveness Score	
DEA Efficiency Score	0.3618***
	(0.1567)
Patient Characteristics	
ESI Level 1 & 2 Patients	2.3798
	(2.1063)
Patients Over 65 Years of Age (%)	1.6429
	(1.9013)
Female Patients (%)	−0.4475
	(2.3294)
White Patients (%)	−1.5158
	(3.5544)
Physician Characteristics	
Physician Experience	−0.0057*
	(0.0025)
Physician Contact-to-Disposition	0.0016*
	(0.0007)
Avg MRI Count per Patient Visit	4.0121
	(3.1387)
Avg IV Count per Patient Visit	−0.0026
	(0.1163)
Avg CT Scan Count per Patient Visit	0.2445
	(0.3874)
Avg Total ED Patients per Shift	−0.0024
	(0.0018)
Observations: $n = 24$	
Note: *p<0.05; **p<0.01; ***p<0.001	

Table 2: Regression Results - Efficiency Model - Individual Physician

Dependent variable: DEA Efficiency Score	
Patient Characteristics	
ESI Level 1 & 2 Patients	-2.5637
	(2.3786)
Patients Over 65 Years of Age (%)	1.8061
	(2.1200)
Female Patients (%)	-1.7110
	(2.1270)
White Patients (%)	3.2178
	(4.6979)
Physician Characteristics	
Experience	0.0035
	(0.0033)
Avg IV order Count	-0.0343
	(0.1185)
Avg Radiology order Count	0.7178
	(0.4133)
Avg MRI Order Count	-6.9205^{*}
	(3.5250)
Avg Total ED Patients per Shift	-0.0020
	(0.0032)
Observations: $n = 24$	
Note: $^{\cdot}$p<0.1; *p<0.05; **p<0.01; ***p<0.001	

With regards to peer effects, our results (presented in Table 3) show a statistically significant positive relationship between a physician's effectiveness and the presence of a more effective peer. This finding suggests that, all else equal, boosting a physician's effectiveness requires scheduling him/her during the same shift as a more effective physician. Similarly, the regression results of peer physician efficiency analysis (displayed in Table 4) show that the presence of a more effective peer is associated with (on average) an increase in physician efficiency. This suggests that effective providers can have a positive influence on their peers' efficiency and effectiveness. Hospital administrators can potentially make use of these findings by incorporating them in their staffing and scheduling decisions[1].

Table 3: Regression Results - Effectiveness Model - Peer Physician

Dependent variable: DEA Effectiveness Score	
Peer Characteristics	
More Efficient Peer	0.0069
	(0.2535)
More Effective Peer	0.0126^{*}
	(0.0411)
More Experienced Peer	-0.0117
	(0.0373)
Different Degree Peer	0.0041
	(0.5593)
Opposite Sex Peer	-0.0098
	(0.2118)
Observations: $n = 522$	
Note: $^{\cdot}$p<0.1; *p<0.05; **p<0.01; ***p<0.001	

[1]It is important to note, however, that our observations may not warrant causal claims. Nevertheless, the natural experiment setting discussed earlier removes potential biases, and hence, many of the associations we find may be causal.

Table 4: Regression Results - Efficiency Model - Peer Physician

Dependent variable: DEA Efficiency Score	
Peer Characteristics	
More Efficient Peer	0.0059
	(0.0081)
More Effective Peer	0.0234**
	(0.0082)
More Experienced Peer	-0.0062
	(0.0075)
Different Degree Peer	0.0232
	(0.0093)
Opposite Sex Peer	-0.0105
	(0.0105)
Observations: $n = 522$	
Note: ·p<0.1;*p<0.05; **p<0.01; ***p<0.001	

Conclusions

Using evidence from emergency medicine, we develop and analyze metrics for physicians' effectiveness and efficiency. Unlike what the conventional wisdom suggests, our findings show that a physician's effectiveness is positively (and not negatively) associated with his/her efficiency. In addition, we find that more effective physicians have higher average contact-to-disposition time and are relatively more recently graduated. We also find that highly efficient physicians have on average lower MRI orders per patient visit. Furthermore, our results provide evidence for the existence of peer effects, and suggest that the presence of a more effective peer has a positive effect on a physician's effectiveness and efficiency.

We believe that our analysis serves as an early step to explore issues of physician effectiveness and efficiency. Importantly, we do not believe that the scores we develop are the only ways to measure effectiveness or efficiency.[2] That is, our work does not provide a definitive calculus for determining who is (or is not) an effective or efficient physician, but rather uses analytical techniques to explore these issues in an early attempt to better understand them. Nevertheless, our findings are valuable because they shed light on potential new ways to improve the efficiency and effectiveness of healthcare delivery.

In particular, our study highlights the need for further theoretical research on the drivers of physicians' performance. Future work can provide a more complete picture of the channels through which a physician's effectiveness and efficiency can be improved. Finally, while our analyses in this paper are based on quantitative data, the inclusion of qualitative factors in future studies may improve the strength and applicability of our effectiveness and efficiency models.

References

Abualenain J, Frohna WJ, Smith M, Pipkin M, Webb C, Milzman D, Pines JM (2013) The prevalence of quality issues and adverse outcomes among 72-hour return admissions in the emergency department. *The Journal of Emergency Medicine*. 45:281-8.

Banker RD, Charnes A, Cooper WW (1984) Some models for estimating technical and scale inefficiencies in data envelopment analysis. *Management Science*. 30:1078-1092.

Charnes A, Cooper WW, Rhodes E (1978) Measuring the efficiency of decision making units. *European Journal of Operational Research*. 2(6):429-444.

Chilingerian JA (1995) Evaluating physician efficiency in hospitals: A multivariate analysis of best practices. *European Journal of Operational Research*. 80:548-574.

Cooper WW, Seiford LM, Tone K (2007) *Data Envelopment Analysis A Comprehensive Text with Models, Applications, References and DEA-Solver Software.*, 2nd ed. (Springer, New York).

Freedman DB (2015) Towards Better Test Utilization Strategies to improve physician ordering

[2]For example, one may improve our scores by also including aspects of patient satisfaction that correlate with higher provider performance levels.

and their impact on patient outcomes. *The Journal of the International Federation of Clinical Chemistry and Laboratory Medicine*. 26(1):15-30.

Gordon JA, An LC, Hayward RA, Williams BC (1998) Initial emergency department diagnosis and return visits: risk versus perception. *Annals of Emergency Medicine*. 32:569-73.

Greene WH (1993) *Econometric Analysis* 2nd ed. (Macmillan, NewYork).

Hand DJ, Till RJ (2001) A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems. *Machine Learning*. 45(2):171-186.

Keehan SP, Stone DA, Poisal JA, Cuckler GA, Sisko AM, Smith SD, Madison AJ, Wolfe CJ, Lizonitz JM (2017) National health expenditure projections, 2016-25: price increases, aging push sector to 20 percent of economy. *Health Affairs (Millwood)*. 36(3):553-563.

Keith KD, Bocka JJ, Kobernick MS, Krome RL, Ross MA (1989) Emergency department revisits. *Annals of Emergency Medicine*. 18:964-8.

Klasco RS, Wolfe RE, Wong M (2015) Assessing the rates of error and adverse events in the ED. *The American Journal of Emergency Medicine*. 33:1786-9.

Pham JC, Kirsch TD, Hill PM, DeRuggerio K, Hoffmann B (2011) Seventytwo-hour returns may not be a good indicator of safety in the emergency department: a national study. *Academic Emergency Medicine*. 18:390-7.

Rousseuw PJ (1987) Silhouettes: a Graphical Aid to the Interpretation and validation of cluster analysis. *Computational and Applied Mathematics*. 20:53-65.

Salway RJ, Valenzuela R, Shoenberger JM, Mallon WK, Viccellio A (2017) Emergency Department (ED) Overcrowding: Evidence-Based Answers To Frequently Asked Questions. *Revista Medica Clinica Las Condes*. 28(2):213-219.

Simar L, Wilson P (1998) Sensitivity analysis of efficiency scores: How to bootstrap in non-parametric frontier models. *Management Science*. 44:49-61.

Simar L, Wilson P (2002) Non-parametric tests of returns to scale. *European Journal of Operational Research*. 139(1):115-132.

Simar L, Wilson P (2011) Inference by the m out of n bootstrap in nonparametric frontier models. *Journal of Productivity Analysis*. 36:33-53.

Traub SJ, Stewart CF, Didehban R, Bartley AC, Saghafian S, Smith VD, Silvers SM, LeCheminant R, Lipinski CA (2016) Emergency department rotational patient assignment. *Annals of Emergency Medicine*. 67(2):206-15.

Tsugawa Y, Ashish KJ, Newhouse J (2017) Variation in physician spending and association with patient outcomes. *JAMA Internal Medicine*. 177(5):675-682.

Wagner JM, Shimshak DG, Novak MA (2003) Advances in physician profiling: the use of DEA. *Socio-Economic Planning Sciences*. 37(2):141-163.