QUALITY-OF-CARE FACTORS THAT DIFFERENTIATE HOSPITALS IN THE NORTHEAST VS OTHER REGIONS

BUDT 733 – SPRING 2008

GROUP 3:

P. KENYON CROWLEY
SELIM GERON
SHIH-HENG HUNG
SETH KELLY
ELIZABETH SLOBASKY

UNIVERSITY OF MARYLAND
ROBERT H. SMITH
SCHOOL OF BUSINESS
Leaders for the Digital Economy

MAY 12, 2008

EXECUTIVE SUMMARY

Hospitals in the United States provide a wide variety of diagnostic and therapeutic medical services in a number of geographic settings. While primarily focused on inpatient treatment, hospitals also provide a range of outpatient, diagnostic x-ray, and pharmacy services. In the last few decades, there has been relatively fast growth in outpatient services and a reduction in inpatient hospital stays, mainly due to regulatory and technological changes as well as increased pressure from managed care companies and governmental Medicare programs.

The goal of this analysis was to look at the characteristics of U.S. hospitals to see if hospitals in certain geographical regions shared common features, particularly in the Northeast versus other areas of the country. The data used was taken from the HIMSS Analytics Database, which comprises data from over 5,000 hospitals. Specifically, we focused on variables relating to finances, quality of care and service metrics, and types of technology used by the facility.

After exploring the data to look at the interaction of the variables, we used classification tree, logistic regression, and discriminant analysis models to explain if a hospital was located in the Northeast or not based on its characteristics. The analysis showed that hospitals in the Northeast tend to be larger hospitals with more electronic medical record (EMR) and computerized physician order entry (CPOE) technologies and are less frequently categorized as critical access type hospitals. The results of the analysis are consistent with our group's expectations, given that these hospitals are located in more populated areas and larger hospitals tend to need—and have the budget for—more sophisticated information technology (IT) systems.

The information gathered from our analysis could have several real-world implications. For example, companies wishing to enter the healthcare IT market as developers of new products may choose to compete in less penetrated markets (i.e., non-northeast locations), while those wishing to support existing infrastructures may choose to compete in the Northeast. Patients seeking out larger, more technologically advanced facilities may choose hospitals in the Northeast more frequently. Depending on their level of comfort with technology in the hospital setting, physicians seeking employment may target more technologically advanced hospitals in their job searches. Similarly, hospital administrators participating in recruiting efforts may highlight certain characteristics of their hospitals compared with others in the United States to increase candidate interest.

TECHNICAL SUMMARY

BACKGROUND AND DATASET

We used the HIMSS Analytics Database, a comprehensive, web-based database containing current and detailed demographic and IT profile information for more than 5000 hospitals nationwide. We chose the healthcare field because of its central relevance to the daily lives of every man, woman, and child. In addition, there is much attention in the media and government given to the issues surrounding healthcare; this project gave our team an opportunity to understand these issues more deeply. The current analysis involves looking at hospitals in one geographical region, the northeastern United States, to contrast them with hospitals in other U.S. locations.

DATA AND ANALYSIS

The first step to understanding the different characteristics between northeast and non-northeast hospitals was to identify the types of hospitals that should be analyzed and clean the data. The analysis focused on four types of hospitals: academic, critical access, general medical and surgical, and general medical. Specialty hospitals or long-term acute hospitals were excluded to avoid confusing the data with hospitals that have alternative missions.

There were quite a few missing values for the revenue and expense variables used in the analysis. There were 1858 missing revenue numbers and 472 missing expense numbers. One option identified to remedy this situation was the use of a regression analysis based on the number of staffed beds for revenue and the number of licensed beds for expenses. Given the large number of missing values, the imputing of that information proved to be inaccurate. A better option was to create dummy variables that indicated whether a hospital reported revenue (*Revenue (Y/N)*), expense, and adjusted patient day (*APD*) information. This revealed whether the act of reporting itself had informative value. Additionally, correlation was calculated between variables to see if variables with missing values could be captured in other variables. Appendix I illustrated the high correlation between *APD*, *Revenue*, and *Staffed Beds*, which essentially can be used as a proxy for hospital size. *Log(Staffed Beds)* was created because *Staffed Beds*, when graphed on a box plot, appeared to be skewed. This remedied the problem and captured the information well.

Pivot tables were created to see if hospital location—as defined by northeast or non-northeast—was a differentiator when displayed versus IT systems type; hospital type; and existence of revenue, expense, and APD reporting. *Revenue y/n*, *CPOE y/n*, *EMR y/n*, and *Service Type_Critical y/n* were the variables that showed the most differentiation (Appendix II).

MODELS

Classification tree and logistic regression models were run on the clean data, both using all variables as well as only the variables found to be helpful in the pivot tables. The tree resulted in all non-northeast nodes and similarly logistic regression classified all hospitals as non-northeast. Since this result is equivalent to the naïve rule, the models were not shown to add value. There were too many non-northeast data points in the dataset used for the analysis, which in turn muted all signals within the northeast hospitals.

Oversampling of northeast hospitals fixed this problem. 2580 non-northeast hospitals were randomly deleted, which left the dataset to include 1/3 northeast and 2/3 non-northeast hospitals. Models run on this dataset were much more informative and were able to unlock the differences between northeast and non-northeast hospitals.

INTERPRETATION

The cutoff value was chosen 0.5 in all models. The classification tree (Appendix III), which we only used for exploratory purposes, indicated that size—as determined by staffed beds—was the only factor. However we still ran a logistic regression (Appendix IV) and discriminant analysis (Appendix V) using the "qualified" variables. A parsimonious logistic regression model with fewer variables (after dropping the insignificant ones) performed slightly better. The model implied the following:

- Every additional staffed bed increases the odds of a hospital being located in the Northeast by a factor of 1.15, for those with same revenue reporting and CPOE status.
- Reporting the revenue increases the odds of a hospital being located in the Northeast by a factor of 2.45, for those with same number of staffed beds and CPOE status.
- Having a CPOE in place increases the odds of a hospital being located in the Northeast by a factor of 2, for those with same number of staffed beds and same revenue reporting status.

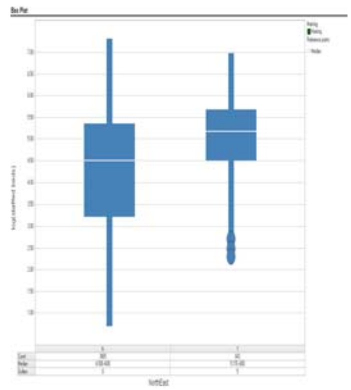
CONCLUSIONS AND RECOMMENDATIONS

Northeast hospitals tend to be larger hospitals with a higher rate of EMR and CPOE adoption and less frequently categorized as critical access type hospitals. This makes sense, given that these hospitals are located in more populated areas and larger hospitals tend to require—and have the budget for— more sophisticated IT systems.

This profile information can be used in several ways. For example, companies considering entering the healthcare IT market with new products may choose to compete in areas where IT currently has less penetration to increase market share, in this case hospitals outside the Northeast. In contrast, companies seeking to offer support for existing products may choose to enter the market where healthcare IT is in relatively widespread use. Patients with the ability to determine where to receive care may choose larger, more resource-laden facilities, like those in the Northeast, which are generally believed to provide better service. Finally, this information may influence physicians' or administrators' decisions on where to seek employment, or could be used as a recruiting tool to showcase how technologically advanced the institution is as compared to others.

APPENDIX I:

	Before Outliers Removed	After Outliers Removed
Adjusted Patient Days and Staffed Bed	0.4012	0.8153
Staffed Beds and Licensed Beds	0.9684	0.9684
Adjusted Patient Days and Revenue	0.2836	0.7266
Revenue and Staffed beds	0.8079	0.8078



APPENDIX II:

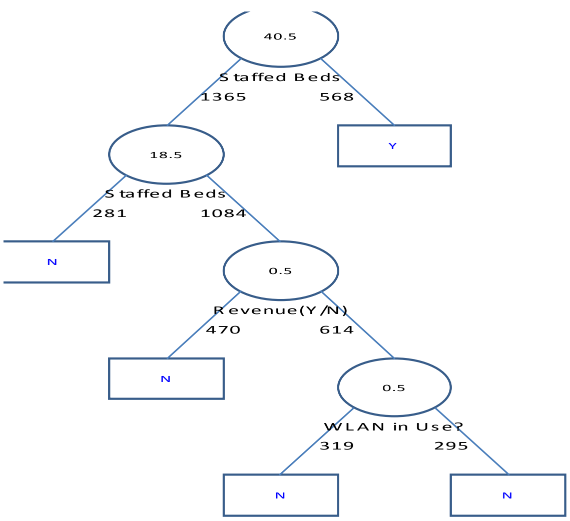
Count of Revenue(Y/N)_Yes	Revenue(Y/N)_Yes		
NorthEast	0	1	Grand Total
N	43.20%	56.80%	100.00%
Y	29.24%	70.76%	100.00%
Grand Total	41.21%	58.79%	100.00%

Count of EMR(Y/N)_Yes	EMR(Y/N)_Yes		
NorthEast	0	1	Grand Total
N	64.90%	35.10%	100.00%
Y	51.48%	48.52%	100.00%
Grand Total	62.99%	37.01%	100.00%

Count of CPOE(Y/N)_Yes	CPOE(Y/N)_Yes		
NorthEast	0	1	Grand Total
N	84.40%	15.60%	100.00%
Y	70.14%	29.86%	100.00%
Grand Total	82.37%	17.63%	100.00%

Count of Service Type_Critical Access	Service Type_Critical Access		
NorthEast	0	1	Grand Total
N	71.73%	28.27%	100.00%
Y	90.36%	9.64%	100.00%
Grand Total	74.38%	25.62%	100.00%

APPENDIX III:



Appendix IV:

The Regression Model

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	-7.12299156	0.4419992	0	*
Staffed Beds	0.14228086	0.01225449	0	1.15290046
Revenue(Y/N)_Yes	0.89739084	0.24928385	0.00031837	2.4531939
CPOE(Y/N)_Yes	0.69752812	0.3088156	0.02390077	2.00878119

Residual df	1929
Residual Dev.	578.5404663
% Success in training data	33.00569064
# Iterations used	12
Multiple R-squared	0.76404309

Training Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable)	0.5

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Y	N
Y	562	76
N	1	1294

Error Report			
Class	# Cases	# Errors	% Error
Y	638	76	11.91
N	1295	1	0.08
Overall	1933	77	3.98

Appendix V:

Classification Function

Variables	Classification Function		
	Y	N	
Constant	-5.11835098	-4.06834269	
Service Type_Critical Access	3.29624462	6.18775129	-2.89150667
Std Log(staffed beds)	4.4949193	-0.54337823	5.03829753
Revenue(Y/N)_Yes	2.98743773	1.80696476	1.18047297
CPOE(Y/N)_Yes	1.03031409	0.85336459	0.1769495
EMR(Y/N)_Yes	1.62822962	1.40668046	0.22154916

Training Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable)	0.5

Classification Confusion Matrix		
	Predicted Class	
Actual Class	Y	N
Y	559	79
N	1	1294

Error Report			
Class	# Cases	# Errors	% Error
Y	638	79	12.38
N	1295	1	0.08
Overall	1933	80	4.14

BIBLIOGRAPHY

1. “IBIS World Industry Report: General Medical and Surgical Hospitals in the US: 62211”, 12/12/2007.
2. Hospital Use of Information Technology: American Hospital Association. http://www.aha.org/aha_app/issues/HIT/index.jsp.
3. “Patients Fare Better In Hospitals Using Information Technology”. ScienceDaily. 12/17/2007. http://www.sciencedaily.com- /releases/2007/12/071215211834.htm
4. “From Charity and Tax Breaks to Healthy Profits”. WSJOnline. 4/8/2008. http://online.wsj.com/article/SB120728196193989017.html.