The Downtown Review

Volume 5 | Issue 1 Article 11

2018

Multiple Regression Models to Analyze Length of Hospitalization due to Nosocomial Infections in U.S. Hospitals

Ernest M. Oleksy
Cleveland State University

Nosocomial diseases are a serious concern and detriment to hospitals' abilities to provide appropriate patient care. Bearing this in mind, an effort must be made to determine whether infection surveillance and control programs have reduced the rates of nosocomial infection at U.S. hospitals. The data that were studied were obtained as part of the **S**tudy on the **E**fficacy of **N**osocomial **I**nfection **C**ontrol (SENIC) and used to develop a model of what variables most contribute to a patient's length of stay at the Cleveland Clinic (Quade et al, 1980).

Studies have been done that analyze how patient length of stay is explained by the contributions of amount of infection staff control (Van der Broeck et al, 2007). Also, length of stay has been studied in medical–surgical intensive care units in Argentina (Rosenthal et al, 2003). Lastly, interhospital analyses have been conducted that studied how nosocomial infections harmed patients with relation to additional charges and extended hospitalization (Haley, 1981).

A random sample of 113 hospitals from the original N=338 hospitals were studied. Each hospital is given an ID number and is measured on 11 other variables. After studying the variables, the information collected was on length of stay (in days), age (in years), infection risk (in percent), routine culturing ratio (ratio of number of cultures performed to number of patients without signs or symptoms of hospital-acquired infection, times 100) routine chest X-ray ratio (ratio of number of X-rays performed to number of patients without signs or symptoms of pneumonia, times 100), number of beds (average number of beds in hospital during study period), whether the hospital is affiliated with a medical school, geographic region in the U.S., average daily census (average number of patients in hospital per day during study period), number of nurses (average number of full-time equivalent registered and licensed practical nurses during study period), and available facilities and services (percent of 35 potential facilities and services that are provided by the hospital). Three possible multiple regression models are tested to investigate how effective they are in predicting patient length of stay in U.S. hospitals. This was done using an online version of the statistical software package none as SAS Studio (https://odamid.oda.sas.com/SASStudio/main?locale=en_US&zone=GMT-04%253A00). Descriptive statistics on this sample is provided in Table 1. There are no missing data, so the sample size for each variable is 113.

Table 1.
Sample demographics on n=113 hospitals.

Variable	Mean	Std Dev
Length of Stay (days)	9.65	1.91
Age (yrs)	53.23	4.46
Infection risk (%)	4.35	1.34
Routine culturing ratio	15.79	10.23
Routine chest X-ray ratio	81.63	19.36
Number of beds	252.17	192.84
Average daily census	191.37	153.76
Number of nurses (full time staff = 1; part time staff = 0.5)	173.25	139.27
Available facilities and services	43.16	15.2
Region (NE %/ NC%/ S%/ W%)	24.78%/ 28.32%/ 32.74%/ 14.16%	
Medical school affiliation (yes %/ no%)	15.04%/ 84.96%	

Results

Table 2
Regression Model Results for Length of Stay

Length of Stay	Model I Parameter Estimate	Model II Parameter Estimate	Model III Parameter Estimate
Intercept	1.38646	6.46738***	8.23171***
Average Age of Patients (yrs)	0.08371*	N/A	N/A
Beds	N/A	0.00302*	0.00541*
Infection Risk	0.65845***	0.64771***	N/A
Available Facilities and Services (% of 35)	0.02174*	-0.00929	0.01522
Nurses	N/A	N/A	-0.00349
Model Values	Value	Value	Value
Overall F	19.12***	18.78***	7.97***
RMSE	1.56840	1.57322	1.75467
Adjusted R^2	0.3267	0.3226	0.1573

*p < 0.05, **p<0.01, ***p<0.001

Modeling Length of Stay of Patients in Model 1:

Based on this study, there is sufficient evidence to conclude that at least one of the explanatory variables (among age, infection risk, and available facilities and services) is useful in explaining length of stay for all patients at the Cleveland Clinic ($F(3,109)=19.12$; $p < 0.0001$). Further inspection of the model (Table 2) indicates that all three predictor variables are significant predictors of length of stay for all patients at the Cleveland Clinic. The estimate of the *y*-intercept was not significant in this model. This model has a moderate fit as measured by the adjusted R^2 of 32.67% and the root mean square error of 1.57 days. The regression model for length of stay for a patient at a U.S. hospital is:

Length of Stay = 1.39 + 0.08 Average age of patients + 0.66 Infection risk + 0.02 Available facilities and services

Modeling Length of Stay of Patients in Model 2:

Based on this study, there is sufficient evidence to conclude that at least one of the explanatory variables (among number of beds, infection risk, and available facilities and services) is useful in explaining length of stay for all patients at U.S. hospitals ($F(3,109)=18.78$; $p < 0.0001$). Further inspection of the model (Table 2) indicates that number of beds and infection risk are significant predictors of length of stay for all patients at U.S. hospitals. Available facilities and services and number of nurses were not found to be significant predictors. The estimate of the *y*-intercept was significant in this model, meaning that if number of beds, infection risk as a percentage, and available facilities and services were zero, the expected length of stay in days would be 6.47 days. This model has a moderate fit as measured by the adjusted R^2 of 32.26% and the root mean square error of 1.57 days. The regression model for length of stay for a patient at U.S. hospitals is:

Length of Stay = 6.47 + 0.003 Number of beds + 0.65 Infection risk – 0.01 Available facilities and services

Modeling Length of Stay of Patients in Model 3:

Based on this study, there is sufficient evidence to conclude that at least one of the explanatory variables (among number of beds, number of nurses, and available facilities and services) is useful in explaining length of stay for all patients at U.S. hospitals ($F(3,109)=7.97$; $p < 0.0001$). Further inspection of the model (Table 2) indicates that number of beds is a significant predictor of length of stay for all patients in this hospital. Available facilities and services and number of nurses were not found to be significant predictors. The estimate of the *y*-intercept was significant in this model, meaning that if number of beds, number of nurses, and available facilities and services were zero, the expected length of stay in days would be 8.23 days. This model has a moderate fit as measured by the adjusted R^2 of 15.73% and the root mean square error of 1.75 days. The regression model for length of stay for a patient at U.S. hospitals is:

Length of Stay = 8.23 + 0.01 Number of beds + 0.02 Available facilities and services – 0.003 Number of nurses

Conclusions/Recommendations

There is a somewhat useful way to explain a patient's length of stay at U.S. hospitals. The most reliable model from the ones that were studied was Model 1, which included average age of patients, infection risk, and available facilities and services as significant predictors of patient length of stay. The most important factor is infection risk. While patient age and available facilities and services are relatively simple to measure, infection risk tends to be a more complicated calculation which is ultimately a mathematical probability, as opposed to a value based on concrete and easily observed instances like the age of a patient or the percentage available facilities and services. In future research, it might be possible to collect other variables that have been shown to be useful in predicting patient length of stay such as routine chest X-ray ratio or routine culturing ratio.

A more effective model may include infection risk but involve two other predictors, since infection risk was the most significant of the three. Two other predictors, along with infection risk, may be more effective in estimating a patient's length of stay. Ultimately, all analyzed models were not great fits for patient length of stay, so more research needs to be conducted before the hospital enacts any policies based on these findings. Also, it may be useful to group patients into different age groups, like adults, pediatrics, and geriatrics, as well as different groups based on reason for hospitalization, so that specific treatment plans can be developed based on modeling that is more suited for different groups of patients that are fundamentally different from each other.

References

Haley, R. W., Schaberg, D. R., Crossley, K. B., Von Allmen, S. D., & McGowan, J. E. (1981). Extra charges and prolongation of stay attributable to nosocomial infections: a prospective interhospital comparison. *The American journal of medicine*, *70*(1), 51-58.

Haley, R. W., Quade, D., Freeman, H. E., Bennett, J. V., & CDC SENIC PLANNING COMMITFEE. (1980). Study on the efficacy of nosocomial infection control (SENIC Project): summary of study design. *American journal of epidemiology*, *111*(5), 472-485.

Rosenthal, V. D., Guzman, S., & Orellano, P. W. (2003). Nosocomial infections in medical-surgical intensive care units in Argentina: attributable mortality and length of stay. *American journal of infection control*, *31*(5), 291-295.

Van den Broek, P. J., Kluytmans, J. A. J. W., Ummels, L. C., Voss, A., & Vandenbroucke-Grauls, C. M. J. E. (2007). How many infection control staff do we need in hospitals?. *Journal of Hospital Infection*, *65*(2), 108-111.