Research Paper
Impact Factor: 3.567
Peer Reviewed Journal

IJMDRR
E- ISSN –2395-1885
ISSN -2395-1877

DETERMINATION OF HOSPITAL LENGTH OF STAY USING STRUCTURAL EQUATION MODELS

B. Senthil Kumar* **R. Ramakrishnan****
**Assistant Professor of Statistics, Periyar E.V.R. College, Tiruchirappalli.*
***Scientist G, Deputy Director, National Institute of Epidemiology (ICMR), Chennai.*

Abstract
In health care, the hospital managements concentrate on length of stay of the patients to meet their future demands. Nowadays, patient arrival rate for admission in the hospital is on the increase. It leads to unavoidable waiting for bed and hence predicting the LOS becomes pertinent. The prediction of length of stay can help hospitals in achieving benchmark performance levels in profitability, patients care and process efficiency. The objective is to determine the LOS of patients using Structural Equation Model. The data are collected from a private hospital in Tamilnadu, South India. A sample of 408 patients is selected for this study. The research seeks to explore whether the relationship between Hospital length of stay, Ventilator Days and Critical Care Unit length of stay are fully or partially mediated by APACHE II Score. The paper present the construction and validation of Structural Equation Model of a mediated model with the dimensions Ventilator Days, Critical Care Unit length of stay and the mediating parameter is APACHE II Score and the outcome is Hospital length of stay. The mediated model is tested with Bayesian testing and estimation. The various diagnostic plots of the Bayesian estimation confirm the fact that convergence takes place and the normality indicates the absolute fit of the regression model.

Key Words: Hospital Length of Stay, APACHE II Score, Bayesian Structural Equation Models.

1. Introduction
In health care, the hospital managements concentrate on Length Of Stay (LOS) of the patients to meet their future demands. Nowadays, patient arrival rate for admission in the hospital is on the increase. It leads to unavoidable waiting for bed both in private as well as in government hospitals and hence reducing the LOS of the patients and predicting the LOS becomes pertinent.

The importance of APACHE II is that it combines in one summary measure the risk factors of physiologic derangement, age and poor chronic health status. The APACHE-II scoring system consists of three components. Acute physiology score (APS) which is the largest component of the APACHE-II score (A2S) is derived from 12 clinical measurements that are obtained within 24 hours after admission to the ICU. The most abnormal measurement is selected to generate the APS component of the A2S. If any variable has not been measured, it is assigned zero points. The variables are internal temperature, heart rate, mean arterial pressure, respiratory rate, oxygenation, arterial pH, serum sodium, serum potassium, serum creatinin, haematocrit, white blood cells count and Glasgow coma scale. The second component is age adjustment: wherein one to six points are added for patients older than 44 years of age. Third component of APACHE-II is chronic health evaluation. An additional adjustment is made for patients with severe and chronic organ failure involving the heart, lungs, kidneys, liver and immune system. [Knaus et al (1985)].

One of the main concerns in the healthcare area is the measurement of flow of patients into the hospitals and other health care facilities. Medhat Hastem et al (2008) have discussed the predictive capability of A2S in determining mortality among the critically ill surgical population. In a commentary in the International Journal of Epidemiology, Tu (2009) has expressed concern about the scarcity of SEM models in epidemiological research and has urged epidemiologists to use SEM models more frequently.

The prediction of LOS can help hospitals in achieving benchmark performance levels in profitability, patients care and process efficiency. LOS is an important measure of health care utilization and determinant of hospitalization costs. Health care providers and hospital administrator are interested in early and accurate LOS predictions for both economic and administrative reasons. [Tanuja et al (2011)].

In this article an attempt has been made to apply Bayesian SEM to derive a new model for the determination of HLOS using APACHE II score as one of the independent variables. It saves time by avoiding collection of data for non contributing independent variables. Here the researcher considers only 3 independent variables.

2. Materials and Methodologies
The data are collected from a private hospital in Tamilnadu, South India. A sample of 408 patients is selected for this study. The independent variables considered are APACHE II score, Ventilator days (VD) (Number of days the patient was using mechanical ventilation) and CCU LOS. The Hospital LOS (HLOS) is a dependent variable.

IJMDRR
E- ISSN –2395-1885
ISSN -2395-1877

The data collected are analyzed for the entire sample. Analysis is performed with Statistical Package for Social Sciences IBM SPSS (Version 20.0) which included descriptive statistics, correlation analysis and AMOS package for SEM and Bayesian estimation and testing. The relationships among HLOS, mediating impact of APACHE II Score with the ventilator days in CCU and CCU LOS are of interest in this study. Mediation refers to a process or mechanism through which one variable causes variation in another variable.

This modeling sequence stresses the importance of the goodness-of-fit assessment. As a combination of regression and path analysis in SEM, each predictor is used with its associated uncontrolled error and, unlike regression analysis, predictor multi collinearity does not affect the model results. The Root mean square error of approximation, goodness-of-fit indices and the normal fit index are used to evaluate model fit.

3. Results and Discussion

Correlation coefficient of the independent variables APACHE II Score ($r = 0.220$, $p < 0.001$), Ventilator days ($r = 0.909$, $p < 0.001$), CCU LOS ($r = 0.950$, $p < 0.001$) with the dependent variable HLOS show highly significant results.

3.1. Regression Model of the DHLOS Mediated SEM

A mediator hypothesis is formulated as:

H_1: Including the interaction between predictor and A2S will explain more of the variance in HLOS than the direct influence of predictors on their own.

The mediator hypothesis is supported if the interaction path (VD, CCULOS X A2S) are significant. There may also be significant main effects for the predictor and mediator (A2S). Therefore, the research seeks to explore whether the relationship between HLOS, VD and CCULOS is fully or partially mediated by A2S. The following sections present the construction and validation of SEM of DHLOS mediated model with the dimensions VD, CCULOS and the mediating parameter A2S and the outcome HLOS. And also, the mediated DHLOS model is tested with Bayesian testing and estimation.

The hierarchical regression for the model was empirically tested with the AMOS graphics environment. The path diagram for the hypothesized mediated model is given in Figure 1.

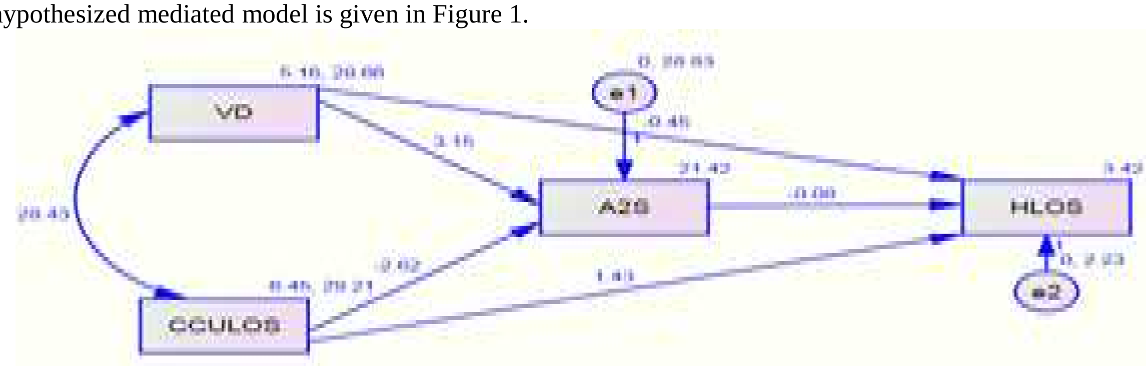


Figure 1: Standardised Parameter Estimates for Mediated DHLOS Model

The analysis is performed for estimating the parameters by using AMOS graphics. Figure 1 shows the standardized parameter estimates.

Table 1: Presents the Maximum Likelihood Estimates

Table 1: Maximum Likelihood Estimates of the Mediated DHLOS Model

Regression Weights	Estimate	SE	C.S	P Value	Label
A2S ← CCULOS	-2.62	0.013	1.001	***	W2
HLOS← CCULOS	1.429	0.003	1.001	***	W4
HLOS ← VD	-0.448	0.003	1.001	***	W5
A2S ← VD	3.152	0.014	1.001	***	W1
HLOS ← A2S	-0.082	0.001	1.001	***	W3
Means					
CCULOS	6.451	0.01	1.001	***	M1
VD	5.158	0.01	1.001	***	M2

IJMDRR
E- ISSN –2395-1885
ISSN -2395-1877

Intercepts					
A2S	21.422	0.012	1	***	M3
HLOS	3.424	0.012	1.001	***	I1
Covariance					
CCULOS<--> VD	28.434	0.082	1.001	***	C1
Variances					
CCULOS	29.208	0.085	1.001	***	V1
VD	28.664	0.081	1.001	***	V2
e2	28.835	0.066	1.001	***	
e1	2.238	0.006	1.001	***	

*** Shows that the p value is significant at 5% level

From the table 1 and the Figure 1 the following points are observed. The regression analysis reveals that CCULOS explained 1.43 of the estimated value of the HLOS, followed by VD which explains 0.45 of the estimated values of the HLOS. The visual representation of results shows the relationships between the dimensions of HLOS and the mediated factor. The VD resulted in a significant impact on the mediated factor, A2S. The CCULOS has caused a negative influence on A2S. It shows that the impact of the factors CCULOS and VD on HLOS is low, where as the impact is very high on the mediating variable. Very high covariance between VD and CCULOS reveals that they play an indispensable role in the outcome of the HLOS.

The above study reveals the relationship among the independent variables and the outcome of the HLOS. The model fit index value of RMSEA and other GFI and NFI are 0.061 (<0.07), 0.987 (>0.95), and 1.000 (>0.95), respectively. Hence, the hypothesized mediated DHLOS model is empirically proved.

3.2. Bayesian Estimation and Testing for Regression Model of DHLOS Mediated SEM

A Bayesian approach was adopted for estimation and inference in AMOS 7.0 environment. The procedure for assessing convergence of Markov Chain Monte Carlo (MCMC) algorithm of maximum likelihood has been adopted. To estimate MCMC convergence, convergence in distribution and convergence of posterior summaries methods were adopted. The values of the posterior means accurately estimate the DHLOS mediated-SEM model. In table 1 the highest value of convergence statistics (C.S) is 1.001 which is less than the 1.002 which is a conservative measure according to Gelman et al (2004).

3.3. Posterior Diagnostic Plots of Regression Model

To check the convergence of the Bayesian MCMC method the posterior diagnostic plots have been analyzed. Figure 2 shows the posterior frequency polygon of the distribution of the parameters across 60,000 samples.

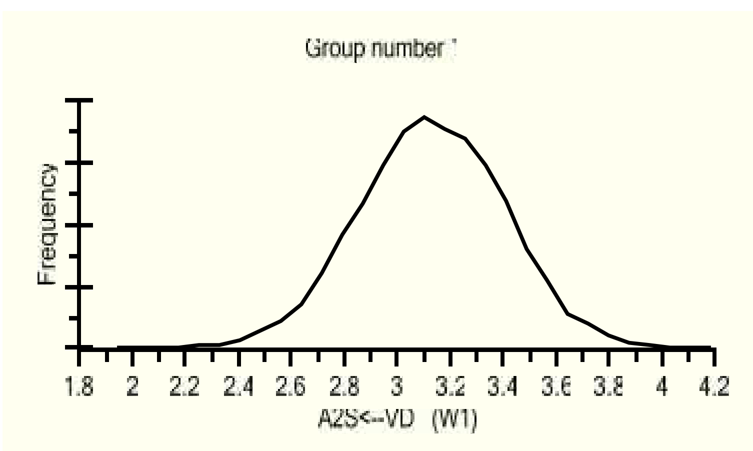


Figure 2: Posterior Frequency Polygon Distribution of the A2S and VD, Regression Weight (W1)

In the same way the posterior frequency polygon distributions of other 4 regression weights were drawn. These Bayesian MCMC diagnostic plots reveal that for all the figures the normality is achieved. Hence the SEM fit is accurately estimated.

IJMDRR
E- ISSN –2395-1885
ISSN -2395-1877

To ensure that AMOS has converged to the posterior distribution, simultaneous displays of two estimates of the distribution are obtained. The first estimate is obtained from the first one third of the accumulated samples and the second is obtained from the last one third. Figure 3 show the simultaneous display of two estimates of the distribution for the mediated factor A2S with CCULOS across 70,000 samples.

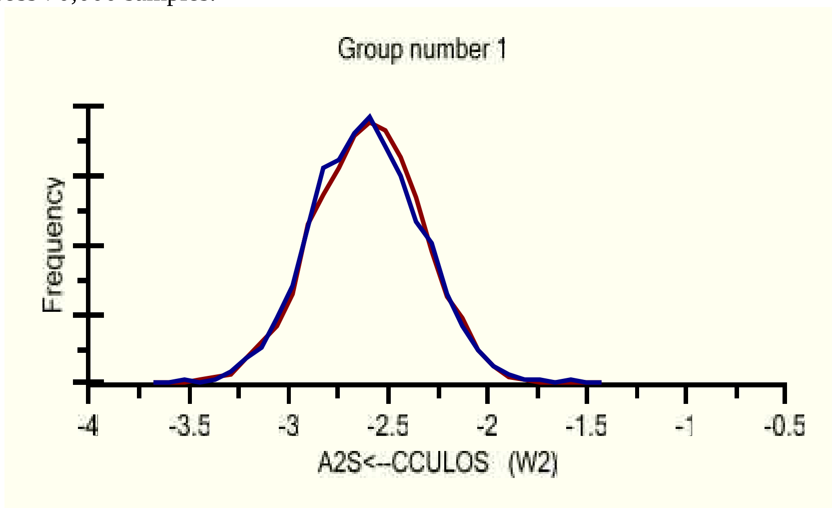


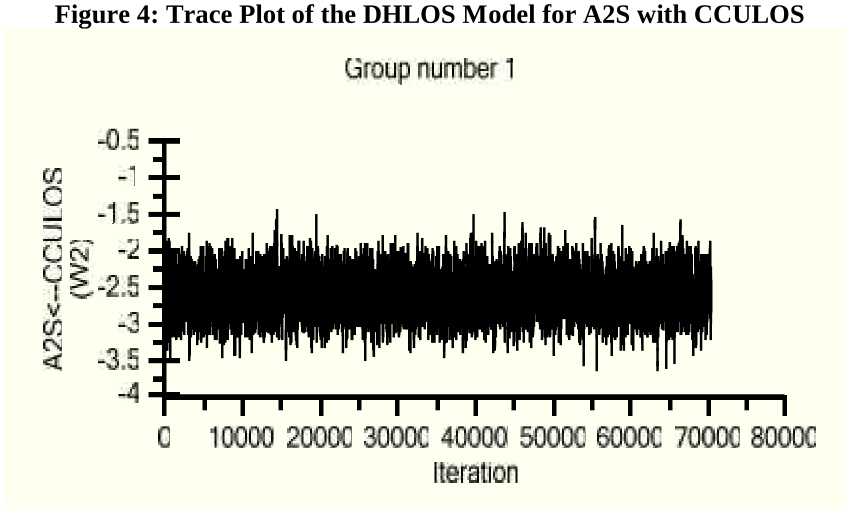
Figure 3: Posterior Frequency Polygon Distributions of the First and Last Third of the Samples of the Regression Model for the Mediated Factor A2S and CCULOS

From the Figure 3 it is observed that the distributions of the first and last thirds of the samples are almost identical. This suggests that AMOS has successfully identified the important features of the posterior distribution of the relationship between the mediated factor A2S and other variables.

3.4. MCMC Convergence in Distribution Using Trace Plot

The trace plot shows the sampled values of a parameter over time. This plot helps us to estimate how quickly the MCMC procedure converges in distribution. Figure 4 show the trace plot of the DHLOS model for the mediated factor A2S with CCULOS across 70,000 samples.

Figure 4: Trace Plot of the DHLOS Model for A2S with CCULOS



The figure 4 exhibits rapid up and down variation with no long term trends. If this plot is broken into a few horizontal sections, the trace within any section would not look much different from the trace in any other section. This result indicates that the convergence in distribution takes place rapidly. Hence, the DHLOS MCMC procedure very quickly forgets its starting values.

3.5. Autocorrelation of DHLOS Model

To determine how long it takes for the correlations among the samples to die down, autocorrelation plot which is the estimated correlation between the samples value at any iteration and the sampled value after k iterations for k=1,2,3,… is analysed for the DHLOS regression model.

Figure 5 shows the correlation plot of the DHLOS model for the mediated factor A2S with CCULOS across 70,000 samples.

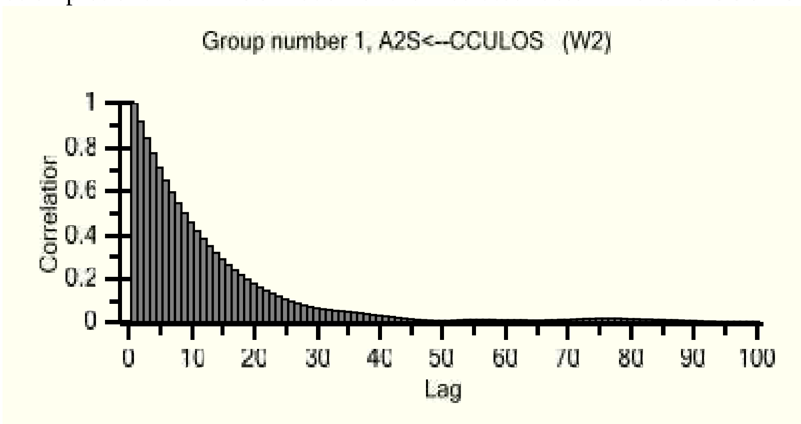


Figure 5: Posterior Correlation Plot of the DHLOS Model for A2S with CCULOS

Figure 5 exhibits that at the lag 50, the correlation is nearer to 0.01. This indicates that by 50 iterations, the MCMC procedure has essentially forgotten its starting position which is equivalent to convergence in distribution. Hence, it ensures that convergence in distribution is attained, and that the analysis samples are indeed samples from the true posterior distribution.

4. Conclusion

The various diagnostic plots featured from the figures of the Bayesian estimation of convergence of MCMC algorithm confirm the fact that the convergence takes place and the normality is attained which indicates the absolute fit of the DHLOS regression model. From the DHLOS regression model which is empirically tested with mediating factor APACHE II score with the dimensions CCULOS, Ventilator days and HLOS, it is evident that the hospital management should concentrate on the APACHE II Score as a mandatory aspect of critical care while this is not given due importance in most of the hospitals in India.

References

1. Knaus, W.A., Draper, E.A., Wagner, D.P., and Zimmerman, J.E., (1985). APACHE II: A severity of disease classification system. Critical care medicine, 13(10), 818-829.
2. Medhat Hashem, Hanan Kamal, Naser Fadel, Walid Awad, Rehab Sami and Ashraf Almasry, (2008). The Predictive Capability of Apache II Score in Determining Mortality among Critically Ill Surgical Population. EJCTA, 2(2), 152 -157.
3. Tu, Y.K., (2009). Commentary: Is structural equation modelling a step forward for epidemiologists. Int J Epidemiol. 38, 1-3.
4. Tanuja, S., Dinesh Acharya., Shailesh, K.R., (2011). Comparison of different data mining techniques to predict hospital length of stay. Journal of pharmaceutical and biomedical sciences (JPBMS), 7 (15), 1-4.
5. Gelman, A., Carlin, J.B., Stern, H.S. and Rubin,D.B. (2004), Bayesian Data.
6. Analysis, 2nd Edition, Chapman and Hall/CRC, Boca Raton, FL.