Distribution Agreement

Signature:

______________________________ ______________

Xin Sun Date

Interrupted time series: a comparison of spine surgery and hip replacement outcomes between a general hospital and a specialty orthopaedic center

By

Xin Sun

Master of Science in Public Health

Biostatistics

Traci Leong, Ph.D.

Committee Chair

Howard Chang, Ph.D.

Committee Member

Interrupted time series: a comparison of spine surgery and hip replacement outcomes between a general hospital and a specialty orthopaedic center

By

Xin Sun

B.E.

Beijing Institute of Technology

2012

Thesis Committee Chair: Traci Leong, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2016

Abstract

Interrupted time series: a comparison of spine surgery and hip replacement outcomes between a general hospital and a specialty orthopaedic center

By Xin Sun

Background: Whether or not a specialty medical center provides better patient outcomes and lower cost compared to a general hospital is controversial. In addition, few studies have focused on the specialty orthopaedic centers. The objective of our study is to evaluate if hospital length of stay (LOS) and cost were improved after the building of an orthopaedic and spine center.

Method: Emory University Orthopaedic & Spine Center (EUOSC) opened in September 2008. Medical records of patients who had spine surgery or hip replacement were collected from Emory University Hospital (EUH) during the period October 2006 to August 2008 and from EUOSC between October 2008 and September 2010. The primary outcome was average LOS and the secondary outcome was average medical cost per hospital day. Interrupted time series analysis with segmented regression model was used for outcomes of spine surgery and hip replacement, respectively. Autocorrelation of time series data was adjusted in the regression model.

Results: Average LOS of spine surgery patients sharply dropped by 0.8 day (change in level, p =.0108) immediately after the opening of EUOSC. In addition, LOS declined by 0.04 hospital day (i.e., one hour) over each month in EUOSC compared to baseline increasing trend in EUH (p=.0345). However, the LOS of patients with hip replacement was not significantly affected by EUOSC (change in level: p=.1448 and change in trend p=.1131). The mean cost per day of spine surgery patients increased by $25 (95%CI: 3.4-46.6) over each month in EUH and once surgeries were conducted at EUOSC this average increased by $1,094 (95%CI: 662.8-1,525.2). After Sept. 2008, the cost decreased by $41 every month in EUOSC compare to EUH (p=.0053). For hip replacements patients, cost had been continually increasing for the four year study period and was not impacted by EUOSC.

Conclusion: For spine surgery patients, LOS decreased immediately when treated at a specialty medical center, although the average cost were higher. For hip replacement patients, there was no evidence of a decreased LOS or increased cost impacted by the specialty orthopaedic and spine center.

Interrupted time series: a comparison of spine surgery and hip replacement outcomes between a general hospital and a specialty orthopaedic center

By

Xin Sun

B.E.

Beijing Institute of Technology

2012

Thesis Committee Chair: Traci Leong, Ph.D.

An abstract of
A thesis submitted to the Faculty of the
Rollins School of Public Health of Emory University
in partial fulfillment of the requirements for the degree of
Master of Science in Public Health
in Biostatistics
2016

Acknowledgments

I would first like to express my deepest gratitude to my thesis advisor, Dr. Traci Leong. She let me choose a research topic I am interested in, encouraged me to think independently and creatively, trained my problem solving skills and steered me in the right direction. Without her enthusiasm, encouragement and continuous support, this thesis would hardly have been completed. I would also like to acknowledge Dr. Howard Chang as the reader of this thesis. I gratefully thank him for his very valuable edits and comments on this thesis.

I am forever indebted to my parents for giving me the love and freedom. They selflessly encourage me to explore new possibilities in life and seek my own destiny. They support every decision I made, especially the best choice I've ever made–studying abroad in the United States, at Emory. I am so lucky to have such wonderful parents.

Contents

1 Introduction

The recent increase in the existence of specialty hospitals——cardiac, orthopedic, surgical and women's hospitals——has drawn the attention of general hospitals and policy makers [1]. Specialty hospitals are not an entirely new phenomenon since pediatric and psychiatric specialty hospitals have existed for decades [2]. Centers for Medicare & Medicaid Services reported that, overall, specialty hospitals provide good quality of care [1]. For example, complications and mortality rates were lower in cardiac specialty hospitals than in general hospitals, along with lower cost and higher patients' satisfaction rate [3]. Other studies have also shown that patient outcomes are better when treated in a specialty hospital compared to a general hospital [4-6].

Opponents argue that the superiority of specialty hospitals might be on account of "cherry picking" healthier patients which will lead to the better clinical outcomes. One study even provided support for both sides of the debate [7]. They claimed that expense for cardiac care is lower in a specialty hospital, however, specialty hospitals tend to treat healthier patients and have less proportion of patients from emergency departments.

However, some studies didn't find any significant differences in clinical outcomes between specialty and general hospital. For instance, specialized urology hospitals were not associated with lower odds of in-hospital mortality after urologic cancer surgery [8].

Among the debate of the two kinds of hospitals, few studies have focused on orthopaedic specialized hospitals or medical centers. In 2007, Cram and colleges compared specialty orthopaedic hospitals with general hospitals and reported better patient outcomes in the specialty hospital after adjusting for patient characteristics and hospital procedural volume [9]. Only one other investigation had assessed outcomes in specialty orthopaedic hospitals and found improvement in the same outcomes [9]. One study also recommended transferring spinal cord injury patients to a spine trauma center due to lower in-hospital mortality rates than the national trauma registry [10].

In our study, the specialty Emory University Orthopaedic & Spine Center (EUOSC) was built located several miles away from the general Emory University Hospital (EUH) in 2008. We collected patient medical data two years before and after 2008. We used interrupted time series analysis to evaluate the change of clinic outcomes and economic outcomes from general hospital to specialty hospital. Specifically, we analyzed spine surgery and hip replacement operation separately in this study because of the hypothesis that spine and hip patients are not comparable. The primary purpose is to evaluate the change in length of stay and the secondary purpose is to evaluate the change in medical expense associated with EUOSC. To our knowledge, it is the first time interrupted time series is used to compare the specialty medical center to a general hospital.

1.1 Interrupted Time Series Review

Randomized controlled trials are considered the gold standard to determine a cause-effect relationship between the intervention and outcome. However, when it is infeasible or not ethical to conduct a randomized controlled trails, quasi-experimental designs are frequently used [11]. Examples include whether or not the campus crime affects applicants to a university, or if we want to assess the impact of economic depression on unemployment rates. Interrupted time series, one of the strongest quasi-experimental design, is used especially to evaluate the impact of a policy or an intervention program [12, 13]. Data are collected at multiple time intervals before and after an intervention (interruption) in order to detect whether the intervention has changed the underlying secular trend [14]. Similar to randomized trials, interrupted time series studies aim to demonstrate causality between an intervention and an outcome. Interrupted time series analysis is widely used in social science, economics and clinic research [13-15]. For instance, Lopez Bernal et al. used interrupted time series analysis to investigate whether the suicide rates in Spain from 2005 to 2010 were affected by the financial crisis in 2008 [16]. Ma et al. evaluated the impact of cigarette tax increase on Pennsylvania adults' smoking prevalence and asthma hospitalization [17]. Hopewell et al. assessed the effect of an editor's implementation of CONSORT guidelines on the reporting of abstracts in journals using interrupted time series [18].

The interrupted time series design is increasingly used to evaluate the effects of health services and policy intervention [13, 19]. In 2002, Madden et al. used interrupted time series analysis to evaluate the effects of a law that guarantee a 48-hour postpartum hospital stay on newborns' health outcomes [20]. In 2003, Gray et al. evaluated the changes of prescribed drug use and cost before and after a intervention to improve hospital antibiotic prescribing [21]. In 2006, Morgan, Griffiths and Majeed studied whether the regulations aimed at reducing paracetamol poisoning is effective by comparing the mortality rates before and after 1999 [22]. In 2008, Garey et al. compared the impact of a change from vancomycin to cefuroxime on surgical site infection rates in patients with cardiac surgery [23]. In 2013, Niven et al investigated the effect a new implemented noninvasive thermometer on adults' fever incidence in the ICU [24]. In 2013, Hawton et al. applied interrupted time series analysis to assess a United Kingdom legislation which was introduced in 1998 to reduce pack sizes of paracetamol. They examined the quarterly changes of paracetamol poisoning death rates and liver transplant activity from 1993 to 2009 [25].

To apply the methodology of interrupted time series analysis, there are three aspects that should be taken into consideration: autocorrelation, seasonal fluctuations and lagged effects [13] . First, autocorrelation often exists in time series data. That is, outcome at time *t* may be correlated with the outcome at time *t*-1, *t*-2, *t*-3 etc. [12]. Correlation between adjacent time points is termed first-order correlation. Correlation between the current time point and two time points (e.g. months) before or after is the second-order autocorrelation and so forth. In the case of ordinary least-squares regression models, one assumes that error residuals should be independent, normal distributed and of a constant variance. If this assumption is violated, the conventional hypothesis test on the intervention effect coefficients is no longer valid due to the underestimation of the standard errors and overestimation of the significance [26]. Thus, if autocorrelation exists, it should be controlled in the regression model.

Secondly, seasonality should be considered, since most interrupted time series studies have a long study period, usually several years in duration. For example, prescription rate in January of one year may be similar to the prescription rate in the previous January because particular illnesses such as influenza occur seasonally. Thus, at least 24 monthly time points is required to model for seasonality [13].

Third, lagged effects means the effect of an intervention (a policy or a campaign) may take time to appear. For example, it may take two or three months to implement the intervention. The transition periods could be excluded from the analysis or analyzed as a separate segment in the model [27].

1.2 Methods for Interrupted Time Series Analysis

Several regression analytical methods have been developed to account for autocorrelation and seasonal trend. The first one is autoregressive integrated moving average (ARIMA) models [26]. It was designed to correct for dependency of the regression residual errors. For example, a university-based hospital in Colombia developed an educational campaign and created a new structured antibiotic order form in order to improve antibiotic use [28].

An ARIMA model was used to investigate the changes of hospital weekly rate of incorrect prescriptions before and after the intervention [28].

Another method is the segmented regression model. Segmented regression models estimates level and trend in the pre-intervention segment and also estimate the changes in level and trend after the intervention. That is, the level and trend of the pre-intervention segment serve as control for the post-intervention segment in order to measure the impact of an intervention [21]. An abrupt intervention effect could be defined by a change in level while an overtime effect is determined by a change in the slopes of the two segments [13]. One of the greatest strength of segmented regression is the intuitive graphical presentation of the outcomes over time, along with the ability to statistically assess to what extent the intervention changed the outcome of interest [13]. Autocorrelation should be checked and adjusted if needed.

Alternative modeling methods have also reported in the literature. Generalized estimating equations (GEE) model with a correlation structure is appropriate for time series data [19]. But it can be problematic when covariates are missing not at random and also the form of odds ratio is not easily interpretable to general audience [19]. Moreover, a generalized linear model assuming a Poisson distribution was used to study the effect of the ban on coal sales on particulate air pollution and death rates in Dublin [29]. Derde et al. used multilevel Poisson segmented regression model to study the interventions aimed at reducing colonization and transmission of bacteria in care units [30]. The advantages of these model are that they can adjust potential confounding factors in the model compared to segmented regression. However, their regression line is not visualized as a two-segment line and cannot assess the immediate change and change over time affected by the intervention.

1.3 Segmented Regression Model

Segmented time series regression analysis is a statistical comparison of time trends before and after the intervention to identify either immediate change in the level of regression line or a sustained change in the slop of the line. Common segmented regression models fit a least squares regression line and assume a linear relationship between time and the outcome within each segment [13]. The linear segmented regression is specified as following,

$$\mathrm{Y_t} = \beta_0 + \beta_1 \times \mathrm{time_t} + \beta_2 \times \mathrm{Intervention_t} + \beta_3 \times \mathrm{time\ after\ Intervention_t} + \mathrm{e_t}$$

- Y_t is the outcome in the time interval t.
- $time_t$ is a continuous variable indicating the period from start of the observation to time t.
- $Intervention_t$ is an indicator whether the intervention occurred before or after the interrupted time point.
- $time\ after\ Intervention_t$ is a continuous variables counting the number of time intervals after the interrupted time point ($\mathrm{time\ after\ Intervention_t} = 0$ before change point, $\mathrm{time\ after\ Intervention_t}$ ==1, 2, 3, 4…. after).
- β_0 estimates the baseline level (intercept) of the outcome at time zero.
- β_1 estimates the slope of outcome over time before the intervention.

- β_2 estimates the change in level immediately after the intervention.
- β_3 estimates the change in trend (slope) after the intervention compared with the trend (slope) before the intervention. That is, the sum of β_1 and β_3 is the post-intervention slope.

Figure 1 illustrated the graphic segmented regression used in interrupted time series analysis [21].

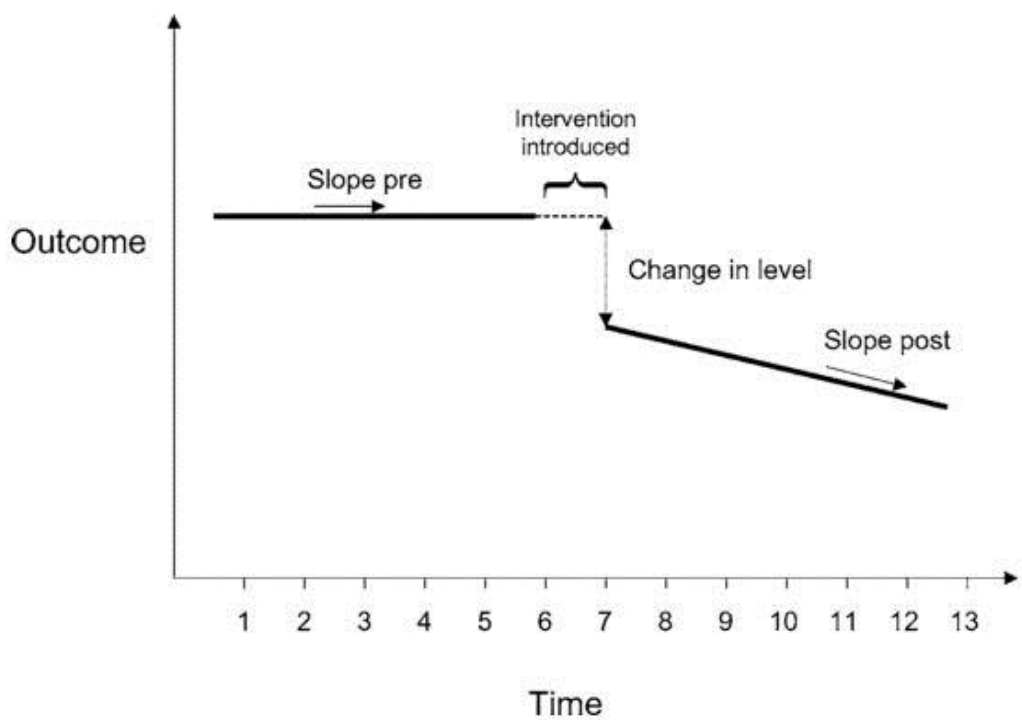

Figure 1. Graphic illustration of interrupted time series using segmented regression

2 Methods

2.1 Setting and Interventions

The setting of our study was Emory University Hospital (EUH) and Emory University Orthopaedic & Spine Center (EUOSC). In September 2008, the department of Orthopaedic & Spine was separated from EUH and relocated three miles away to function entirely as an orthopedics and spine specialty hospital. EUOSC offers refined diagnostic and advanced in-patient surgical procedures, such as hip and knee replacements, spine surgery, bone restoration etc. It is an extension of EUH and the only dedicated orthopaedics and spine surgical facility in Atlanta. By concentrating all orthopaedic surgical services at one location, special needs of the orthopaedic patient population could potentially be better met. In addition, the combination of a dedicated hospital with a university hospital system may offer patients a higher level of specialized care. Thus, this study planned to evaluate whether or not the outcomes (hospital length of stay and cost) truly are better in this specialty unit EUOSC than general hospital EUH.

In our study, medical records were collected from October 2006 to September 2010. EUOSC opened in the middle of September 2008, thus we considered this time to be the interrupted time point. Since the new EUOSC needed some time for transition, the medical data in September 2008 was removed in the regression model and treated as lagged effects [27]. That is, from October 2006 to August 2008, medical records were collected in EUH and from October 2008 to September 2010, medical records were obtained in EUOSC, Thus, there were 23 months of data prior to the change and 24 months of follow up data after this change.

2.2 Study Sample

In this study, we focused separately on the outcomes of spine surgery and hip replacement. For spine operations, there were 3989 patient visits during the study period. Among them, 1,602 medical records obtained from EUH (Oct. 2006 – Aug. 2008) and 2387 records from EUOSC (Oct. 2008 – Sept. 2010). For hip replacement operations, EUH had 743 patient visits from Oct. 2006 to Aug. 2008 and EUOSC got 1161 visits from Oct. 2008 to Sept. 2010 for a total of 1904 patient visits in the study period.

2.3 Outcomes and Demographic Variables

There are two outcomes in our study. The clinical outcome is hospital length of stay (LOS) and the economic outcome is COST which is the estimated cost based on hospital charges. Data were aggregated by month, thus 48 time intervals were generated (23 months pre-intervention, 1 month during the intervention and 24 months post-intervention). The primary outcome LOS was the average LOS aggregated by month, calculated by total patients hospital days divided by the number of patients in that month. The secondary outcome is COST which is the average hospitalization expense per hospital day, calculated by averaging patients' total cost divided by his LOS in that month.

Other patient level data, like age, sex, race, primary payer, risk of mortality, severity of illness, and admission source were provided. There were three payer categories: *Public* included Medicaid and Medicare, *Private* included commercial or private insurance and *Other* consisted of government-assisted health care, military, worker's compensation etc. For admission source, non-health care facility point of origin was considered as *New*, clinic referral, transfer from a hospital, transfer from a skilled nursing facility or intermediate care facility, transfer from another health care facility were grouped as *Referral/Transfer*. Patients came from emergency room were grouped as *Emergency*.

2.4 Statistical Analysis

We compare the patient characteristics by hospital using two-sample t-tests for the continuously distributed variables such as age and Chi-Square tests for categorical variables such as sex and race.

We created the scatter plot of outcomes over time to visually inspect our data and examine whether there was an intervention effect. A two-sample t-test was also applied to outcomes of the pre- and post-time period to test if there was any difference. Then we analyzed the 48-month data as an interrupted time series using segmented regression analysis. Here, the intervention was treated as the creation of a new specialty orthopaedic and spine medical center. In the 48 monthly intervals, there were 23 months prior to the invention, 1 month when the specialty medical center was in transition, and 24 months after the intervention.

Data in that transition month is eliminated in the analysis since we treated that month as a transition period.

Autocorrelation was evaluated through visual detection of the plot of the residuals vs. time; and through the Durbin-Watson test. If the result of Durbin-Watson test is significant, autocorrelated errors in the regression model need to be controlled for. One key assumption of the segmented regression is that residual errors are independent of each other. However, with time series data, the ordinary regression residuals are often correlated over time. A stepwise autoregression method that initially fits a high order model with many autoregressive lags was conducted using PROC AUTOREG in SAS. For example, we could set the lag to be 6 which means the highest order of autocorrelation is 6 time units. Then autoregressive error model could be expressed as

$$y_t = \boldsymbol{X'}_t\boldsymbol{\beta} + V_t$$

where $V_t = -\varphi_1 v_{t-1} - \varphi_2 v_{t-2} - \cdots - \varphi_6 v_{t-6} + \varepsilon_t \qquad \varepsilon_t \sim IN(0, \sigma^2)$

Then by specifying the BACKSTEP option, insignificant autoregressive lags were backward eliminated until all remaining lags have significant t test [18].

We also took seasonal trends into account. In order to detect the seasonality, at least 24 monthly data points are suggested [13]. We have 47 monthly time points which was sufficient to check seasonality. In order to account for seasonal trends, a lag of 12 (ie. one for each month) is often used [31]. Thus, we implemented a segmented regression model using PROC AUTOREG by specifying a lag of 12 and BACKSTEP options to automatically test and estimate autoregressive parameters.

Segmented regression analysis is a powerful method to estimate the level and trend change affected by move of the Orthopaedic & Spine department from EUH to EUOSC. Segmented regression models fit a least squares regression line in each segment and assumes a linear relationship between the outcome and independent variables [13]. The maximum likelihood method was applied to fit the regression models and estimate the standard errors of regression coefficients. The following model was used in our study.

$$\mathrm{Outcome_t} = \beta_0 + \beta_1 \times \mathrm{month_t} + \beta_2 \times \mathrm{Intervention_t} +$$

$$\beta_3 \times \text{month after } \mathrm{Intervention_t} + v_t$$

$$v_t = -\varphi_1 v_{t-1} - \varphi_2 v_{t-2} - \cdots - \varphi_{12} v_{t-12} + \varepsilon_t \qquad \varepsilon_t \sim IN(0, \sigma^2)$$

Where:

- $\boldsymbol{Outcome_t}$ is the primary outcome LOS or secondary outcome COST for spine surgery and hip replacement.
- $\boldsymbol{\beta_0}$ estimates the baseline average patient LOS or COST in October 2006.
- $\boldsymbol{\beta_1}$ estimates the slope change in mean LOS or COST per person that occurs with each month before the intervention in EUH (i.e. **baseline trend**).
- $\boldsymbol{month_t}$ is a continuous variable indicating the number of month since October 2006.

- $\boldsymbol{\beta_2}$ estimates the level/intercept change in mean LOS or COST immediately after moving to EUOSC in September 2008 (i.e. **change in level**).
- ***Intervention*** is a binary indicator for two types of hospital, EUH coded as 0 and EUOSC codes as 1.
- $\boldsymbol{\beta_3}$ estimates the change in trend/slope in the mean LOS or COST in specialty EUOSC compared to the initial slope of EUH (i.e. **change in trend**).
- $\boldsymbol{month\ after\ Intervention_t}$ is a continuous variable which equals to $(\mathrm{month_t}$ -23), indicating the number of months after September 2008 (coded as 0 before September 2008, coded as 1 for October 2008, 2 for November 2008 and so on).
- $\boldsymbol{v_t}$ is an autocorrelated error with an initial highest order of autocorrelation set to 12 to account for autocorrelation and seasonality change.

A p value of less than 0.05 was considered to indicate statistical significance. SAS version 9.4 (SAS Institute, Cary, NC) was used for our data analysis.

3 Results

3.1 Spine Surgery

The baseline characteristics of patients with spine surgery across the two hospitals were presented in Table 1. In the two years before the opening of EUOSC, there were 1,602 patient visits in EUH for spine surgical operation. After EUOSC opened in September 2008, EUOSC treated 2,387 patients with spine surgery in the following two years. Patients' age was not significantly different in two hospitals (mean age: EUH =56.8 vs. EUOSC=56.9 year-old, p=.8506). Nevertheless, sex, race, primary payer, risk of mortality, severity of illness, admission source of spine surgery patients differed in the two hospitals. More male patients underwent spine surgical operations in EUOSC (53.3%) compared to EUH (49.8%), p-value=0.0333. More White patients (77.8% vs. 73.4%) and more Black patients (17.9% vs. 13.1%) went to EUOSC. There was a higher proportion of patients with private health insurance at EUH (58.3%) compared to EUOSC (53.8%). In addition, EUOSC had relatively less mild-risk of mortality patients (73.2% vs. 77.1%), less mild severity of illness patients (35.0% vs. 41.2%) and more moderate and severe patients. Meanwhile, EUOSC had more referral and transfer patients (9% vs. 3.1%) and less emergency patients (2.6% vs. 4.9%).

3.1.1 Length of Stay (LOS)

The mean LOS of patients who had spine surgery at EUH between October 2006 and August 2008 was 5 days while median LOS in EUOSC between October 2008 and September 2010 was 4 days ($p<.0001$). Thus segmented regression was conducted to assess the effect of the opening of EUOSC on spine surgery patients' LOS while controlling for the baseline trend in EUH. Results of segmented regression were presented Figure 2 and Table 2. There was a significant change in level (β_2, p=.0108) and trend (β_3, p=.0345) of LOS after moving to EUOSC. Before the beginning of the study period, average LOS of patients who received a spine surgical operation at EUH was five days. During the first 23 months in EUH, the difference of LOS over time was not significantly (p=.4956). However, after the opening of EUOSC, the average LOS sharply dropped by 0.8 day per month (p =.0108). In addition, monthly change of LOS is significantly decreasing by 0.04 day (i.e., 1 hour) for the next 24 months in EUOSC compared to the trend in EUH (p =.0345). As for autocorrelations, only lag 10 was significantly detected (p=.0333) among all 12 lags and then it was adjusted in the regression model. The Durbin-Watson statistics of the final model after controlling for the autoregression equaled to 2.13 (p-value for testing positive autoregression was .4988 and p-value for testing negative autocorrelation was .5012). Model diagnostic was shown in Appendix Figure 1.

3.1.2 Cost

The secondary outcome is average patients' medical charges per hospital day. The mean COST in general hospital EUH and specialty EUOSC were $5,952 and $7,111 respectively ($p<.0001$). As showed in Figure 3 and Table 3 prior to the opening of EUOSC (September 2008), the average baseline medical expense spent at EUH is $5,640 and increased by $25 every month ($p =.0285$). After September 2008, the cost increased abruptly by $1,094 once EUOSC opened ($p <.0001$). However, during the two years period in EUOSC, the cost had been decreasing by $ 41 per month over time compared to the increasing trend in EUH ($p =.0053$). In other words, after September 2008, the medical charges for patients who underwent spine surgery at EUOSC was decreasing by $16 every month. Lag 7 was detected ($p=.0060$) and adjusted. The Durbin-Watson statistics after controlling for the autoregression was 2.55 (p-value for testing positive autoregression was .9293 and p-value for testing negative autocorrelation was .0707). Model diagnostic was shown in Appendix Figure 2.

3.2 Hip Replacement

As shown in Table 4, the mean age of patients in EUH was 64.8 (±17.9) years old while the mean age in EUOSC was 63.6 (±16.0) (p =.1586). There were no difference in distribution of sex in two hospitals (p=.6381). Payer source were different in two hospitals (p<.0001). More than half patients (59.6% in EUH and 54.2% in EUOSC) were Medicare and Medicaid patients. More patients had private insurance in EUOSC (43.7%) than EUH (35.3%). Risk of mortality was comparable before and after the intervention (p=.4776). However, severity of illness was significantly different between two hospitals. EUOSC has more mild illness patients (33.1% vs. 23.4%) and less severe patients (17.4% vs. 29.4%) (p< .0001). In addition, EUOSC had more transferred patients (7.9% vs. 4.3%) and less emergency patients (15.2% vs. 25.3%) compared to EUH.

3.2.1 LOS

Before the intervention, the mean LOS of patients who underwent hip replacement was 4 days and after the intervention, the mean LOS was 3 days ($p<.0001$). Results of segmented regression were presented in Figure 4 and Table 5. The change of LOS over time was not impacted by the intervention in either the level (β_2, p=.1448) or in the trend (β_3, p=.1131). The average LOS at the beginning of this study was 4 days. During the 23 months in EUH, no significant decrease in average LOS was detected (p=.6781). There was also no remarkable drop of LOS after the opening of EUOSC (p=.1448). Although LOS appeared to be decreasing after intervention (visually), the monthly change was not significantly different than prior to the intervention (p=.1131). No autocorrelation was detected. The Durbin-Watson Statistics was 2.1340 (p-value for testing positive autocorrelation=.5016 and p-value for testing negative autocorrelation=.4984). Model diagnostic was shown in Appendix Figure 3. Appendix Figure 5 and Table 1 showed the segmented regression results of LOS for both spine surgery patients and hip replacement patients.

3.2.2 Cost

The mean medical expense per hospital for hip replacement patients was $4,054 in EUH and the median expense was $5,336 in EUOSC (p<.0001). Medical charges were increasing significantly (by an average of $33 per month) in the pre-intervention period (p-value=.0028) (Figure 5 and Table 6). However, no immediately level change (β_2: p=.1270) and trend change (β_3: p=.5603) of medical charges were detected after the opening of the specialty EUOSC. The slope of post-intervention period is the sum of β_1 and β_3; that is, cost increased by $41 every month after September 2008. Compared to the pre-intervention slope, the post-intervention slope was not significant (p=.5603). Lag 7 was adjusted in this model. The Durbin-Watson Statistics was 1.9226 (p-value for testing positive autocorrelation=.2323 and p-value for testing negative autocorrelation=.7677). Model diagnostic was shown in Appendix Figure 4. Appendix Figure 6 and Table 2 showed the segmented regression results of cost for both spine surgery patients and hip replacement patients.

4 Discussion

Interrupted time series with segmented regression was applied to analyze the level and trend change before and after September, 2008 when the department of the Orthopaedic & Spine at general hospital EUH was transformed to a specialty medical center EUOSC. As expected, LOS of patients receiving spine surgery decreased immediately and over time when treated in specialty unit EUOSC. The better outcomes for spine surgery in specialty hospital may due to larger hospital volume, as higher hospital volume is reported to have positive association with patient outcomes [32].

Unfortunately, we did not observe significant decrease of LOS impacted by specialty EUOSC with hip replacement operations. The first reason may be because of the smaller sample size for hip replacement data. It is suggested that the time points of interrupted time series should be larger than 24, while the observations for each time points that used to calculated the mean should be more than 100 in order to diminish the variation [13]. However, there are approximately 30-60 observations aggregated in a month which may cause large variation of time series data. We considered a sensitivity analysis by aggregating the outcomes by 2 months. The results still didn't show evidence of significant changes in level (β_2: p=.3882) and trend (β_3: p=.2344) of LOS. By looking at Figure 4, we could see that the trend of LOS was decreasing over time, although not statistically significant.

Cost (the average medical expense per hospital day) for both spine surgery and hip replacement patients was higher in specialty EUOSC than general EUH. This result contradicted the statements of those advocates of specialty hospital, that, specialty hospitals lower cost [7]. One possibility may be a result of the more severe patients at EUOSC. While there were more severe spine surgery patients, there were less severe hip replacement patients in EUOSC. Other possibilities might be the advanced equipment, high hospitalization fee or medical inflation. However, we could not ignore the significantly decreasing trend of spine surgery patients' cost following the opening of EUOSC. In contrast, the medical charges for hip replacement patients was significantly increasing over the four years and was not influenced by the intervention.

There are several strengths to our study. First, to our knowledge, this is the first study to evaluate the difference between general and specialty medical facilities focusing on orthopaedics and spine surgery using interrupted time series analysis. Secondly, interrupted time series analysis has several advantages over other quasi-experimental studies, such as controlling for baseline trends. If the outcomes is increasing or decreasing before an intervention is implemented, then data need to be adjusted for these trends to estimate true intervention effect [15]. Thirdly, one great advantage of segmented regression is the visualized graphic presentation which is easy to understand for those without a statistical background. In addition, autocorrelation was detected and adjusted in the model avoiding under- or overestimates of intervention effects. Another advantage of segmented regression is the statistical assessment of level and trend change influenced by the intervention, compare to other common regression models for interrupted time series such as ARIMA models. ARIMA models are often used in the analysis of autocorrelated time series data, but requires at least 50 consecutive time points [15]. Fourth, segmented regression model gives more information than two sample t test in the before-and-after study design. On one hand, most times outcomes are not normal distributed, although we could use central limit theorem is sample size is large enough. On the other hand, we could control for the baseline trend of pre-intervention period and also assess the level and trend change using segmented regression.

Nevertheless, limitations also exist in our study. First, a linear trend was assumed in the outcome with each segment in our segmented regression model. However, the length of stay and cost may have non-linear trend over time. If non-linear patterns are detected, ARIMA models would be more appropriate [28]. Secondly, small number of observations for each time interval can cause large variation, as observed in the hip replacement data. For data with insufficient observations in one month, we could expend the study period and aggregate data quarterly. Thirdly, causal inference from interrupted time-series design is limited, since it is impossible to rule out other unmeasured or uncontrolled factors that might have influenced changes in LOS and COST. To overcome this, a reference outcome could be brought in [16, 33]. In order to verify the impact of financial crisis on suicide rates, Lopez Bernal et al. also analyzed the change of death rates caused by accidental falls over time as a reference [16]. The reference outcome should be similar to our study outcomes, but unlikely to be affected by intervention. In our study, we could find a comparable general hospital similar to EUH. But most times it is difficult to find a suitable control group. Fourth, we could not adjust the demographic variables such as sex, race, payer source and severity of illness in the segmented regression model. Because the outcomes are aggregated in a month to get 47 time points of average LOS and cost, the demographic variable such as sex and race could not be aggregated into month intervals. In the future analysis, we could stratify by these groups and conduct segmented regression respectively if sample size is large enough.

Reference

1. Dummit, L.A. (2005). Specialy hospitals: can general hospiutal compete? *National Health Policy Forum*, (804), 1-12.
2. GAO. (2003). Specialty hospitals: geographic location, services provided and financial performance. United States General Accounting Office.
3. Leavitt, M.O. (2005). Study of physician-owned specialty hospitals required in section 507(c)(2) of the medicare prescription drug, improvement, and modernization act of 2003. Centers for Medicare & Medicaid Services.
4. Sah, B.K., et al. (2009). Effect of surgical work volume on postoperative complication: superiority of specialized center in gastric cancer treatment. *Langenbecks Arch Surg, 394*(1), 41-47.
5. Vernooij, F., et al. (2008). Specialized care and survival of ovarian cancer patients in the Netherlands: nationwide cohort study. *J Natl Cancer Inst, 100*(6), 399-406.
6. Whisker, L., et al. (2009). Appendicitis in children: a comparative study between a specialist paediatric centre and a district general hospital. *J Pediatr Surg, 44*(2), 362-367.
7. Barro, J.R., R.S. Huckman, and D.P. Kessler. (2006). The effects of cardiac specialty hospitals on the cost and quality of medical care. *J Health Econ, 25*(4), 702-721.
8. Konety, B.R., et al. (2006). Mortality after major surgery for urologic cancers in specialized urology hospitals: are they any better? *J Clin Oncol, 24*(13), 2006-2012.
9. Cram, P., et al. (2007). A comparison of total hip and knee replacement in specialty and general hospitals. *J Bone Joint Surg Am, 89*(8), 1675-1684.
10. Kattail, D., J.C. Furlan, and M.G. Fehlings. (2009). Epidemiology and clinical outcomes of acute spine trauma and spinal cord injury: experience from a specialized spine trauma center in Canada in comparison with a large national registry. *J Trauma, 67*(5), 936-943.
11. Harris, A.D., et al. (2006). The use and interpretation of quasi-experimental studies in medical informatics. *J Am Med Inform Assoc, 13*(1), 16-23.
12. Biglan, A., D. Ary, and A.C. Wagenaar. (2000). The value of interrupted time-series experiments for community intervention research. *Prevention Science, 1*(1), 31-49.
13. Wagner, A.K., et al. (2002). Segmented regression analysis of interrupted time series studies in medication use research. *Journal of Clinical Pharamacy Therapeutics, 27*, 299-309.
14. Ramsay, C.R., et al. (2003). Interrupted time series designs in health technology assessment: lesson from two systematic reviews of behaviour change strategies. *International Journal of Technology Assessment in Health Care, 19*(4), 613-623.
15. Matowe, L.K., et al. (2003). Interrupted Time Series Analysis in Clinical Research. *The Annals of Pharmacotherapy, 37*(7), 1110-1116.
16. Lopez Bernal, J.A., et al. (2013). The effect of the late 2000s financial crisis on suicides in Spain: an interrupted time-series analysis. *Eur J Public Health, 23*(5), 732-736.

17. Ma, Z.Q., et al. (2013). Use of interrupted time-series method to evaluate the impact of cigarette excise tax increases in Pennsylvania, 2000-2009. *Prev Chronic Dis, 10*, E169.
18. Hopewell, S., et al. (2012). Effect of editors' implementation of CONSORT guidelines on the reporting of abstracts in high impact medical journals: interrupted time series analysis. *BMJ, 344*, e4178.
19. Zhang, F., et al. (2009). Methods for estimating confidence intervals in interrupted time series analyses of health interventions. *J Clin Epidemiol, 62*(2), 143-148.
20. Madden, J.M., et al. (2002). Effects of a law against early postpartum discharge on newborn follow-up, adverse events, and HMO expenditures. *N Engl J Med, 347*(25), 2031-2038.
21. Ansari, F., et al. (2003). Outcomes of an intervention to improve hospital antibiotic prescribing: interrupted time series with segmented regression analysis. *J Antimicrob Chemother*, *52*(5), 842-848.
22. Morgan, O.W., C. Griffiths, and A. Majeed. (2007). Interrupted Time-Series Analysis of regulations to reduce paracetamol (acetaminophen) poisoning. *PLoS Med*, *4*(4), e105.
23. Garey, K.W., et al. (2008). Interrupted time series analysis of vancomycin compared to cefuroxime for surgical prophylaxis in patients undergoing cardiac surgery. *Antimicrobial Agents and Chemotherapy, 52*(2), 446-451.
24. Niven, D.J., et al. (2013). Fever in adult ICUs: an interrupted time series analysis. *Crit Care Med, 41*(8), 1863-1869.
25. Hawton, K., et al. (2013). Long term effect of reduced pack sizes of paracetamol on poisoning deaths and liver transplant activity in England and Wales: interrupted time series analyses. *BMJ, 346*, f403.
26. Euitema, B.E. and J.W. McKean. (2007). An improved portmanteau test for autocorrelated errors in interrupted time-series regression models. *Behavior Research Methods*, *39*(3), 343-349.
27. Feldstein, A.C., et al. (2006). Reducing warfarin medication interactions: an interrupted time series evaluation. *Arch Intern Med, 166*, 1009-1015.
28. Perez, A., et al. (2003). An interrupted time series analysis of parenteral antibiotic use in Colombia. *Journal of Clinical Epidemiology, 56*(10), 1013-1020.
29. Clancy, L., et al. (2002). Effect of air-pollution control on death rates in Dublin, Ireland: an intervention study. *The Lancet*, *360*, 1210-1214.
30. Derde, L.P.G., et al. (2014). Interventions to reduce colonisation and transmission of antimicrobial-resistant bacteria in intensive care units: an interrupted time series study and cluster randomised trial. *Lancet Infectious Diseases, 14*(1), 31-39.
31. Penfold, R.B. and F. Zhang. (2013). Use of interrupted time series analysis in evaluating health care quality improvements. *Acad Pediatr*, *13*(6 Suppl), S38-44.
32. Hughes, R.G., S.S. Hunt, and H.S. Luft. (1987). Effects of surgeon volume and hospital volume on quality of care in hospitals. *Medical Care, 25*(6), 489-503.
33. Fowler, S., et al. (2007). Successful use of feedback to improve antibiotic prescribing and reduce Clostridium difficile infection: a controlled interrupted time series. *J Antimicrob Chemother, 59*(5), 990-995.

4.1 Tables and Figures

Table 1 Demographic characteristics of patients with spine surgery

	EUH (General) N=1,602	**EUOSC (Specialty)** N=2,387	**p-value**
Age (years)	56.8 (±15.2)	56.9 (±14.7)	0.8506
Sex (Male)	798 (49.8)	1,271 (53.3)	0.0333
Race			
White	1,176 (73.4)	1,856 (77.8)	<0.0001
Black	209 (13.1)	428 (17.9)	
Others	217 (14.5)	103 (4.3)	
Primary Payer			
Private	934 (58.3)	1,284 (53.8)	0.0093
Public	562 (35.1)	903 (37.8)	
Others	106 (6.6)	200 (8.4)	
Risk of Mortality			
Mild	1,235 (77.1)	1,600 (73.2)	0.0039
Moderate	256 (16.0)	373 (17.1)	
Severe	111 (6.9)	213 (9.7)	
Severity of Illness			
Mild	660 (41.2)	766 (35.0)	<0.0001
Moderate	623 (38.9)	859 (39.3)	
Severe	319 (19.9)	561 (25.7)	
Admission Source			
New	1473 (92.0)	2110 (88.4)	<0.0001
Referral/Transfer	50 (3.1)	214 (9.0)	
Emergency	79 (4.9)	63 (2.6)	

EUH represents Emory University Hospital while EUOSC represents Emory University Orthopaedic & Spine Center. Continuous variable age is reported as mean (SD). P-value is given by two sample t test for age and Chi-Square test for categorical data.

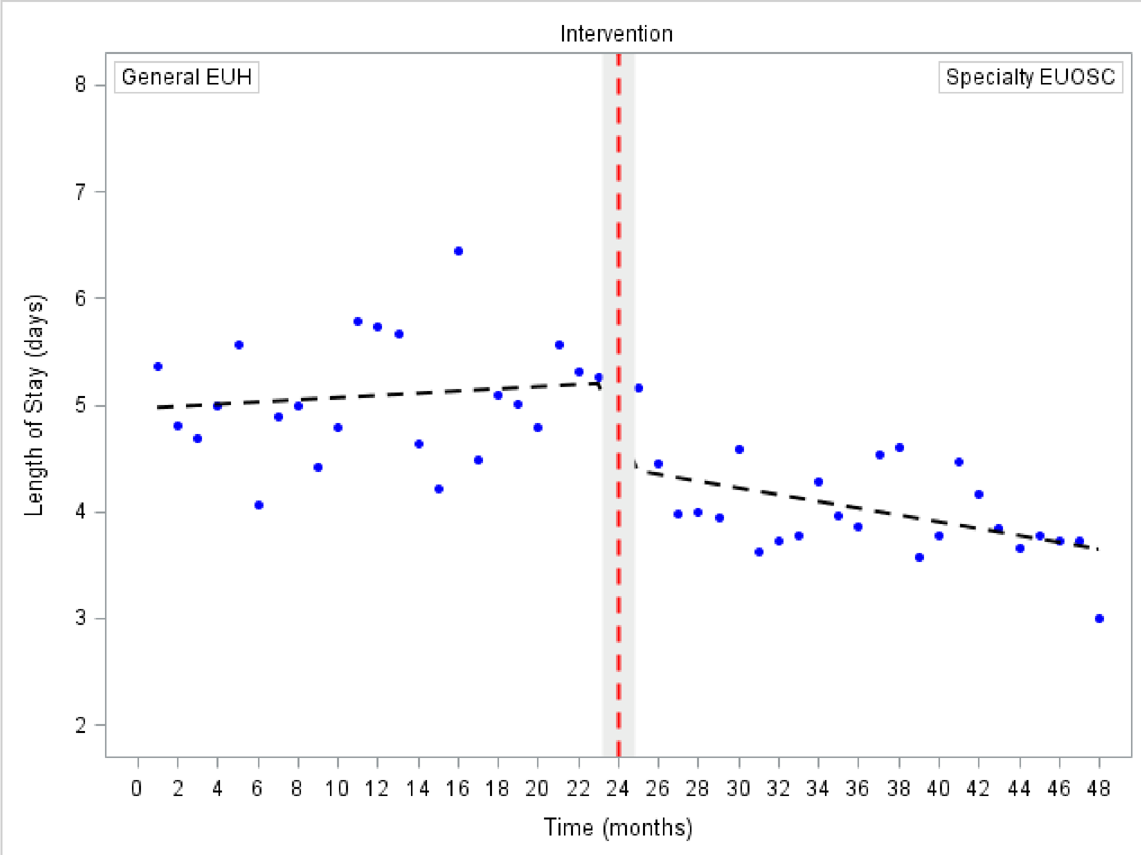

Figure 2 Effects of EUOSC on average LOS of patients who had spine surgery from Oct. 2006 to Oct. 2010 using segmented regression

Table 2 Results of segmented regression model predicting LOS of spine surgery patients

Variable	Coefficient	Standard Error	p-value
Intercept (β_0)	4.96	0.20	<0.0001
Baseline trend (β_1)	0.01	0.02	0.4956
Change in level after the intervention (β_2)	-0.80	0.30	0.0108
Change in trend after the intervention (β_3)	-0.04	0.02	0.0345

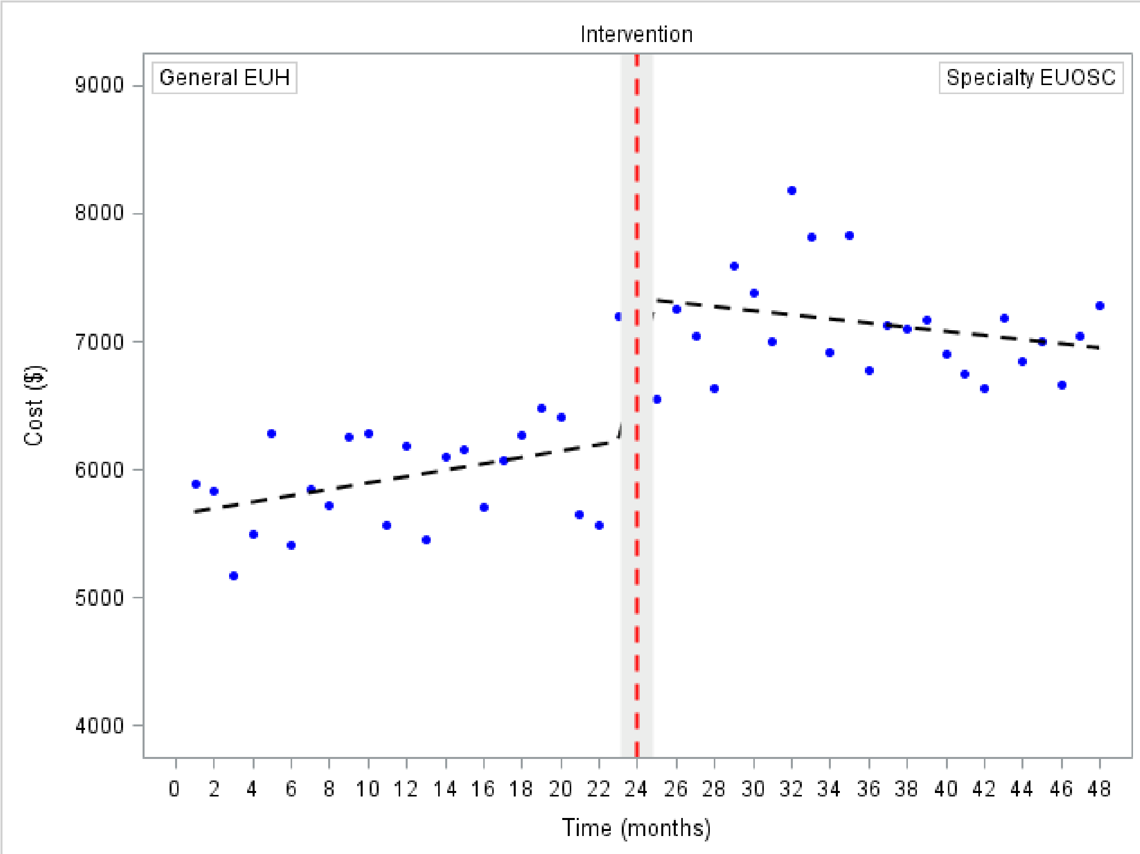

Figure 3 Effects of EUOSC on average COST of patients who had spine surgery from Oct. 2006 to Oct. 2010 using segmented regression

Table 3 Results of segmented regression model predicting COST of spine surgery patients

Variable	Coefficient	Standard Error	p-value
Intercept (β_0)	5640	147	<0.0001
Baseline trend (β_1)	25	11	0.0285
Change in level after the intervention (β_2)	1094	220	<0.0001
Change in trend after the intervention (β_3)	-41	14	0.0053

Table 4 Demographic characteristics of patients with hip replacement

	EUH (General) N=743	**EUOSC (Specialty)** N=1,161	**p-value**
Age (years)	64.8 (±17.9)	63.6 (±16.0)	0.1586
Sex (Male)	305 (41.1)	464 (40.0)	0.6381
Race			
White	510 (68.6)	850 (73.2)	<0.0001
Black	164 (22.1)	267 (23.0)	
Others	69 (9.3)	44 (3.8)	
Primary Payer			
Private	262 (35.3)	507 (43.7)	<0.0001
Public	443 (59.6)	629 (54.2)	
Others	38 (5.1)	25 (2.1)	
Risk of Mortality			
Mild	519 (69.9)	795 (68.6)	0.4776
Moderate	172 (23.1)	265 (22.9)	
Severe	52 (7.0)	99 (8.5)	
Severity of Illness			
Mild	174 (23.4)	384 (33.1)	<0.0001
Moderate	351 (47.2)	573 (49.5)	
Severe	218 (29.4)	202 (17.4)	
Admission Source			
New	523 (70.4)	893 (77.0)	<0.0001
Referral/Transfer	32 (4.3)	92 (7.9)	
Emergency	188 (25.3)	176 (15.2)	

EUH represents Emory University Hospital while EUOSC represents Emory University Orthopaedic & Spine Center. Continuous variable age is reported as mean (SD). P-value is given by two sample t test for age and Chi-Square test for categorical data.

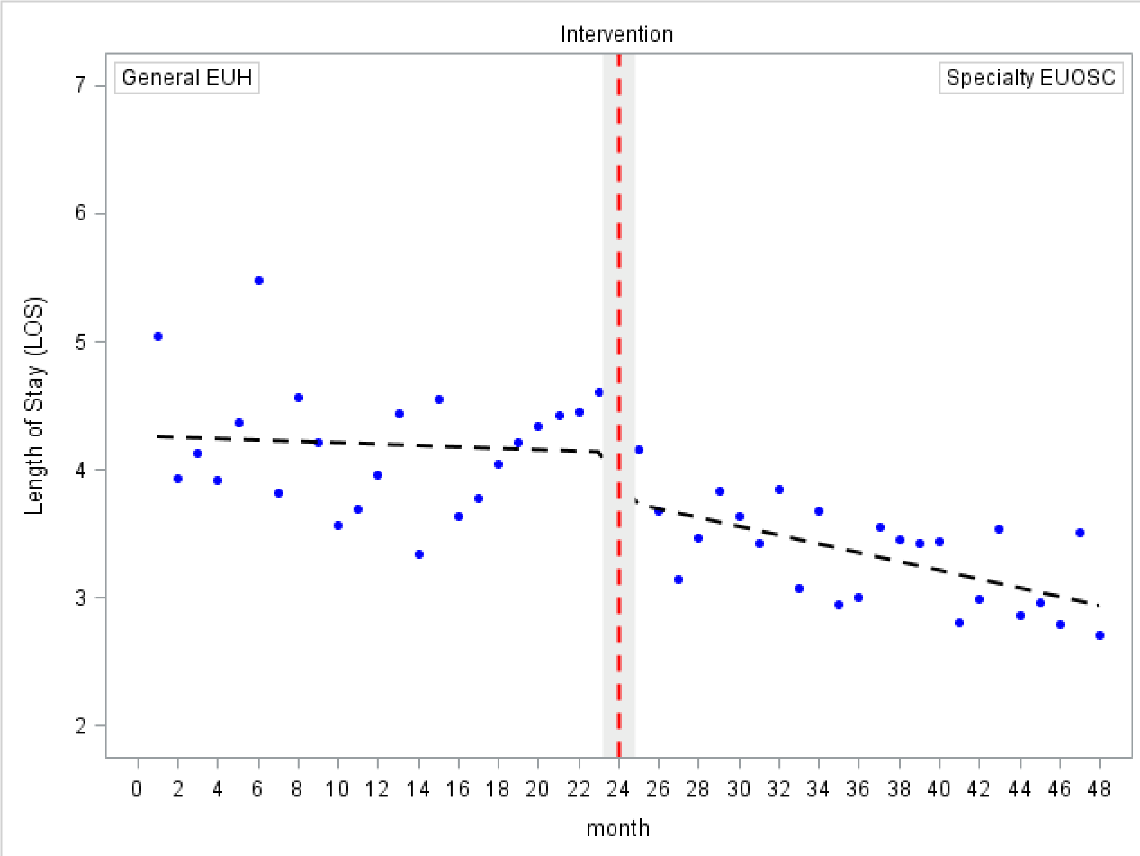

Figure 4 Effects of EUOSC on average LOS of patients who had hip replacement from Oct. 2006 to Oct. 2010 using segmented regression

Table 5 Results of segmented regression model predicting LOS of hip replacement patients

Variable	Coefficient	Standard Error	p-value
Intercept (β_0)	4.26	0.18	<0.0001
Baseline trend (β_1)	-0.005	0.01	0.6781
Change in level after the intervention (β_2)	-0.37	0.25	0.1448
Change in trend after the intervention (β_3)	-0.03	0.02	0.1131

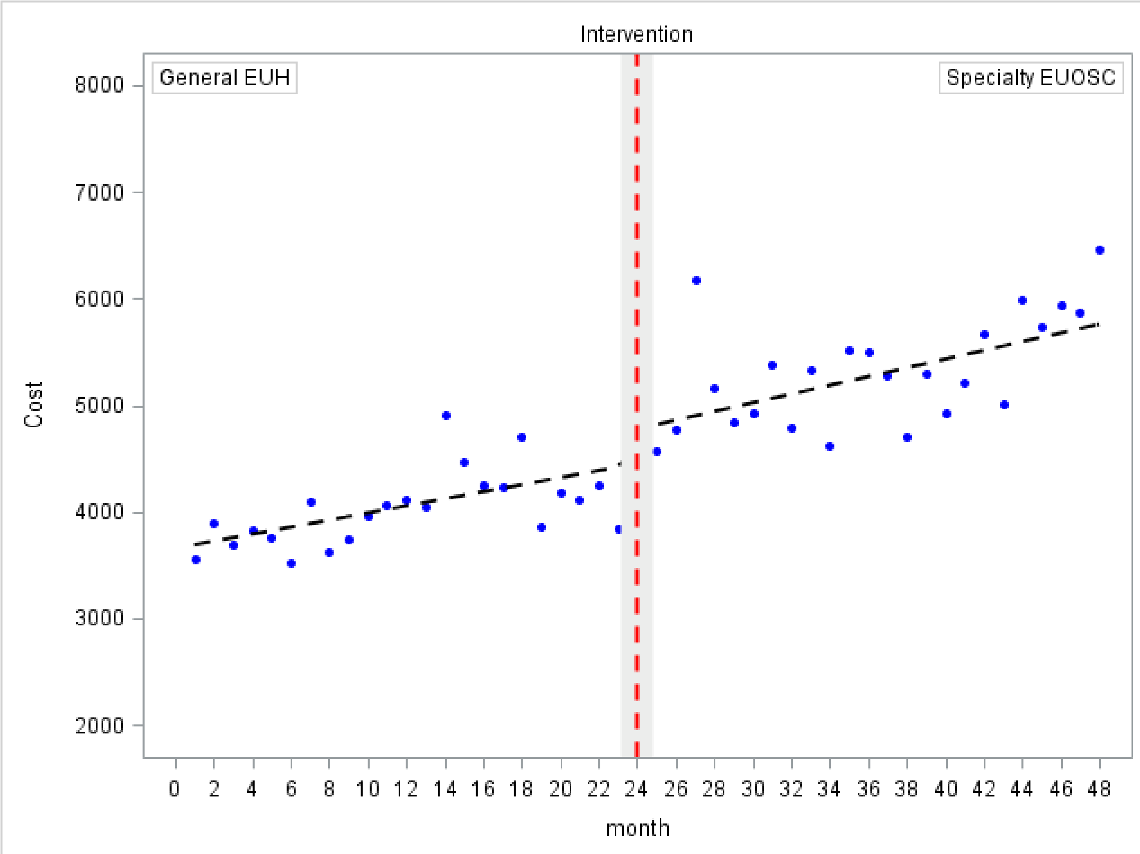

Figure 5 Effects of EUOSC on average COST of patients who had hip replacement from Oct. 2006 to Oct. 2010 using segmented regression

Table 6 Results of segmented regression model predicting COST of hip replacement patients

Variable	Coefficient	Standard Error	p-value
Intercept (β_0)	3656	139	<0.0001
Baseline trend (β_1)	33	10	0.0028
Change in level after the intervention (β_2)	324	208	0.1270
Change in trend after the intervention (β_3)	8	13	0.5603

5 Appendix

5.1 Tables and Figures

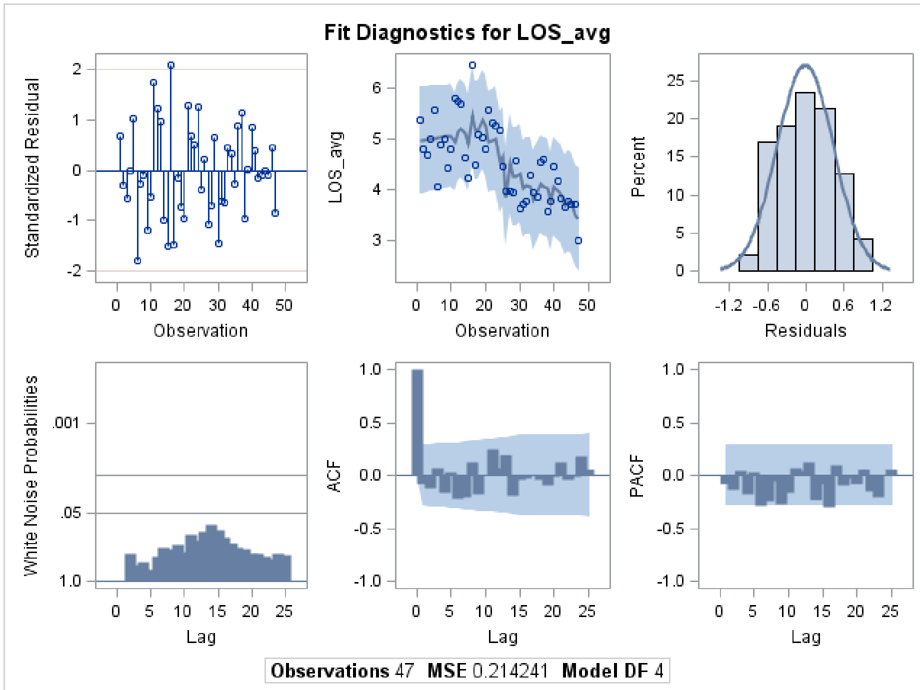

Figure 1 The diagnostics for segmented regression model with outcome LOS of spine surgery patients

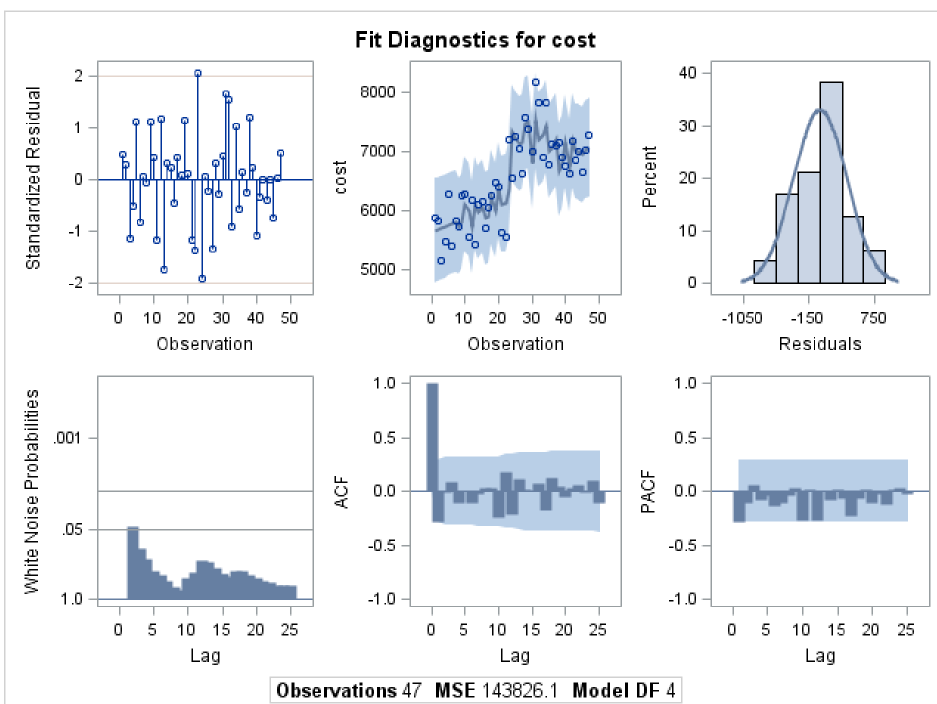

Figure 2 The diagnostics for segmented regression model with outcome COST of spine surgery patients

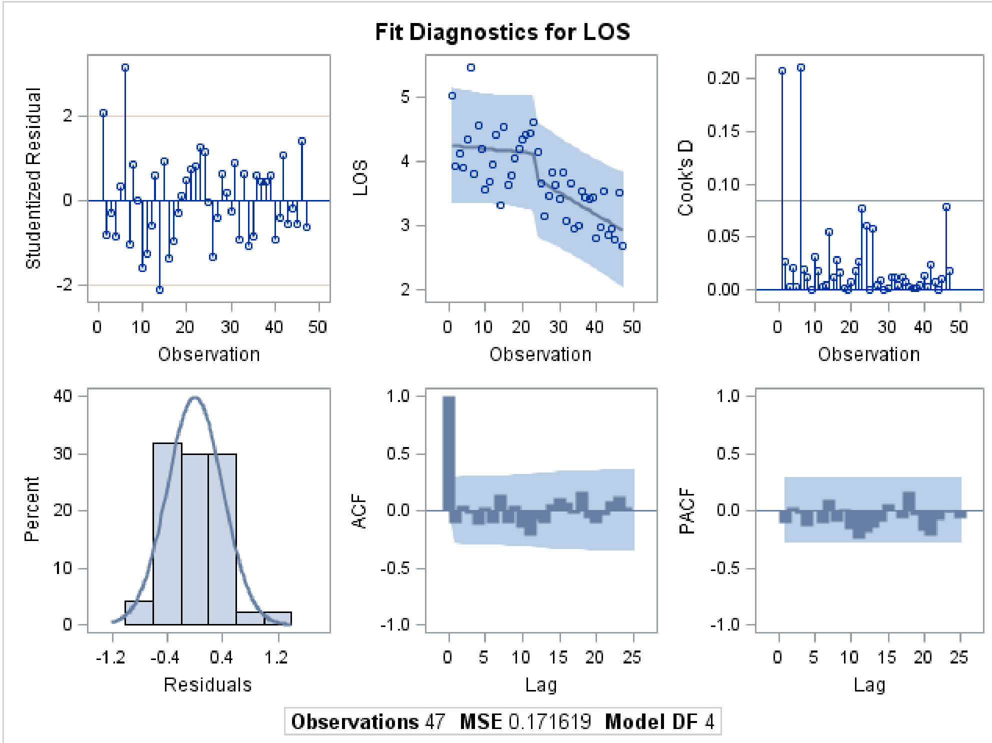

Figure 3 The diagnostics for segmented regression model with outcome LOS of hip replacement patients

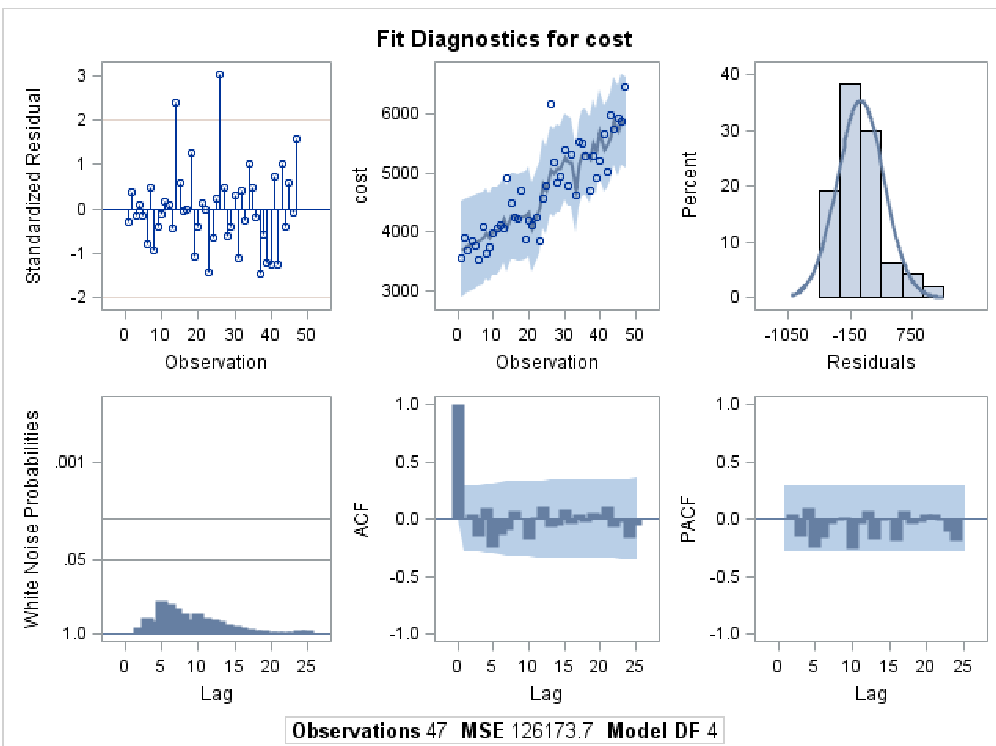

Figure 4 The diagnostics for segmented regression model with outcome COST of hip replacement patients

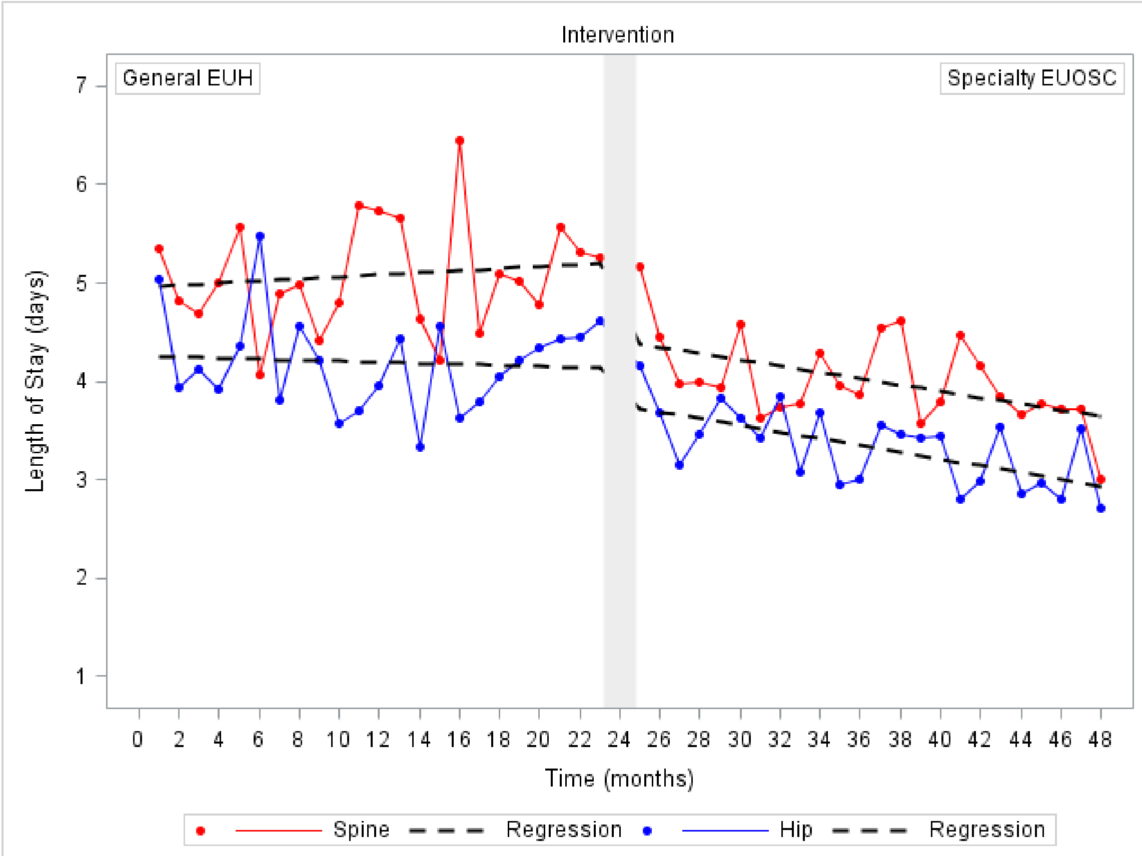

Figure 5 Effects of EUOSC on average LOS of patients who had spine surgery and hip replacement respectively from Oct. 2006 to Oct. 2010 using segmented regression

Table 1 Results of Segmented Regression Model Predicting LOS over time

	Spine Surgery		Hip Replacement	
Variable	Coefficient (SE)	p-value	Coefficient (SE)	p-value
Intercept (β_0)	4.96 (0.20)	<0.0001	4.26 (0.18)	<0.0001
Baseline trend (β_1)	0.01 (0.02)	0.4956	-0.005 (0.01)	0.6781
Chang in level after the intervention (β_2)	-0.80 (0.30)	0.0108	-0.37 (0.25)	0.1448
Change in trend after the intervention (β_3)	-0.04 (0.02)	0.0345	-0.03 (0.02)	0.1131

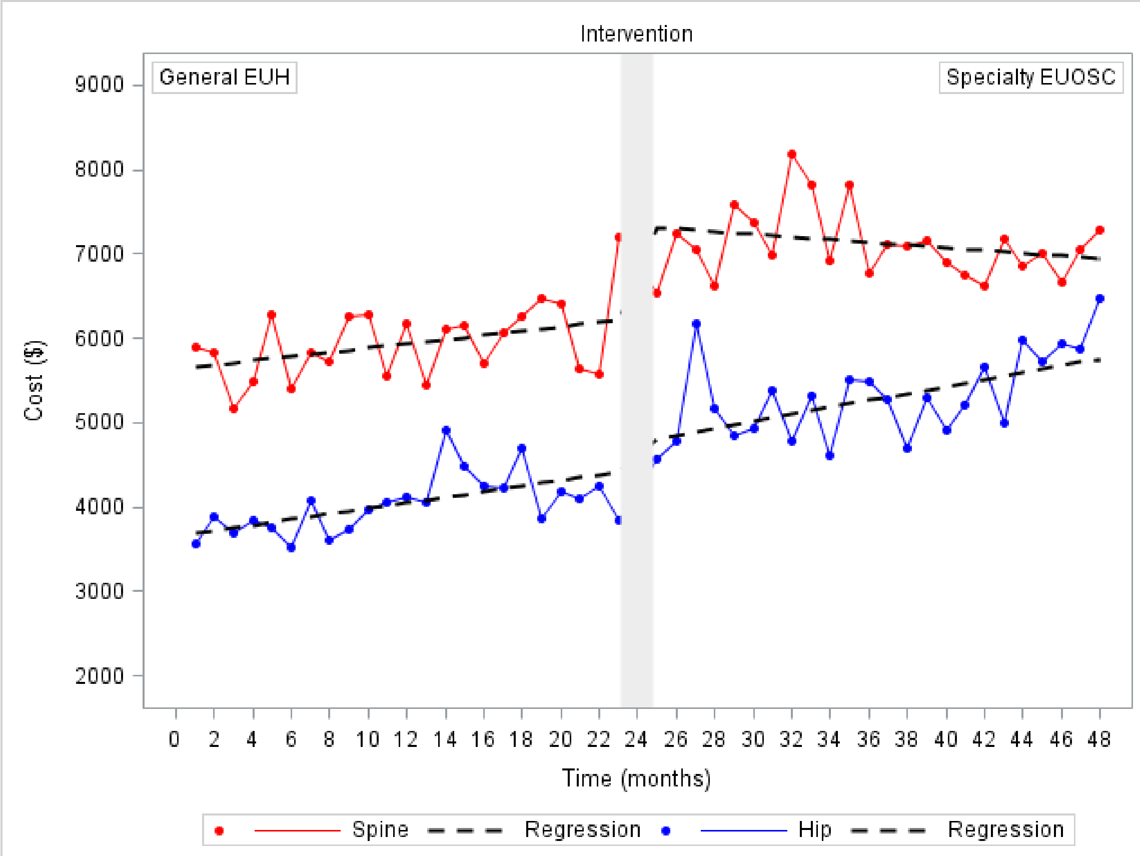

Figure 6 Effects of EUOSC on average COST of patients who had spine surgery and hip replacement respectively from Oct. 2006 to Oct. 2010 using segmented regression

Table 2 Results of Segmented Regression Model Predicting COST over time

	Spine Surgery		Hip Replacement	
Variable	Coefficient (SE)	p-value	Coefficient (SE)	p-value
Intercept (β_0)	5640 (147)	<0.0001	3656 (139)	<0.0001
Baseline trend (β_1)	25 (11)	0.0285	33 (10)	0.0028
change in level after the intervention (β_2)	1094 (220)	<0.0001	324 (208)	0.1270
Change in trend after the intervention (β_3)	-41 (14)	0.0053	8 (13)	0.5603

5.2 SAS code

```
*import spine surgery data;
PROC IMPORT OUT= WORK.spine
            DATAFILE= "H:\Childrens
care\thesis\QQ1_Spine_procedures_for_Traci.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="QQ1_Spine_procedures_for_Traci";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

*import hip surgery  data;
PROC IMPORT OUT= WORK.hip
            DATAFILE= "H:\Childrens
care\thesis\QQ1_Hip_Procedures_for_Traci.xlsx"
            DBMS=EXCEL REPLACE;
     RANGE="QQ1_Hip_Procedures_for_Traci";
     GETNAMES=YES;
     MIXED=NO;
     SCANTEXT=YES;
     USEDATE=YES;
     SCANTIME=YES;
RUN;

PROC FORMAT;
value sex 1="Male" 2="Female";
value race 1="White" 2="Black" 3="Others";
value payer 1="Private" 2="Public" 3="Others";
value rom 1="Mild" 2="Moderate" 3="Major and extreme";
value soi 1="Mild" 2="Moderate" 3="Major and extreme";
value hospital 0="General" 1="Specialty";
value source 1="New" 2="Refereral/Transfer" 3="Emergency";
RUN;

/**********************************SPINE
SURGERY****************************************/

/*****create new dataset for spine surgery*****/
data thesis_spine;
 set spine;    *set hip;
*convert age from month to years;
 age=age/12;
*aggregate into month;
month=intck('month','21sep2006'd,AdmissionDate);   * by 2 month--
month2,by 3 month--qtr;
if month=0 then delete; *we start our study preroid from Oct 1, 2006;
if month=24 then delete; *treat month 24 as transition period.
*define two types of hospital;
if month<24 then hospital=0; else hospital=1;
```

```
*create variable for segmented regression;
if month <24 then month_after=0; else month_after=month-24;

*recode variables;
adm_source=AdmissionSourceCode+0;
if adm_source=1 then source=1;
   else if adm_source in (2,4,5,6) then source=2;
   else source=3;
ID=compress(PatientId,'','kd');
pri_payer=floor(PrimaryPayerCode/100);
if pri_payer=1 then payer=1;
  else if pri_payer in (2,3) then payer=2;
  else payer=3;
*combine groups and make race three groups;
if racecode in (3,6) then race=3; else race=racecode;
*combine groups and make ROM three groups;
risk_mortality=ROM+0;*convert char variable to num;
if risk_mortality in (3,4) then risk=3; else risk=risk_mortality;
*combine groups and make SOI three groups;
if SOI=4 then SOI=3;
*calculate cost per hospital day;
cost=totalcost/LOS;

format sexcode sex. race race. payer payer. risk rom.  soi soi.
hospital hospital. source source. ;
label age="Age (years)";
label sexcode="Sex";
label race="Race";
label payer="Primary payer";
label risk="Risk of Mortality";
label SOI="Severity of Illness";
label source="Admission Source";
run;


/*****compare demographic variables between two types of hospital*****/
*two sample t test ;
proc TTEST data=thesis_spine sides=2 alpha=0.05 h0=0;
 class hospital ; var age LOS COST; run;
*rank sum test for LOS and cost which are skwed distributed;
proc means data=thesis_spine mean std median p25 p75;
class hospital; var LOS cost; run;
proc NPAR1WAY data=thesis_spine wilcoxon;
class hospital; var LOS cost; run;

*chisquare test for categorical variables;
PROC FREQ data=thesis_spine;
tables sexcode*hospital race*hospital payer*hospital risk*hospital
soi*hospital source*hospital/chisq;
run;

*caculate how many observations are aggregated into each month;
proc freq data=thesis_spine;
tables month;
run;

*first look the scatter plot of outcome vs. day---no extrem outliners;
```

```
PROC GPLOT DATA=thesis_spine;
     PLOT LOS*month=hospital ;
RUN;

*caculate average LOS by month;
PROC MEANS data=thesis_spine;
class month; var los;
run;

/*********Time Series************/
*caculate the average LOS and Cost by month and creat dataset called
average;
PROC SQL;
create table average_spine as
 select month as month, month_after as month_after, avg(LOS) as
LOS_spine, hospital as hospital, avg(totalcost/LOS) as COST_spine,
        age as age, sexcode as sex, race as race, payer as payer, risk
as risk, soi as soi, source as source
  from thesis_spine
   group by month;
quit;
*selct the time series points and output as csv file;
PROC SORT data=average_spine; by month; run;
data ITS_spine; set average_spine; by month;
if first.month then output ITS_spine;
*if month=24 then delete;
label LOS_Spine="Length of Stay (days)";
label COST_spine="Cost ($)";
label month="Time (months)";
run;


*export the time series data to csv;
PROC EXPORT data=ITS_spine
    outfile='H:\Childrens care\thesis\time_series_spine.csv'
    dbms=csv
    replace;
run;


/*********segmented regressioon of LOS***********/
*scatter plot of aggregate time series outcome (48ponits) vs. month;
PROC GPLOT DATA=ITS_spine;
     PLOT LOS_Spine*month=hospital COST_spine*month=hospital;
RUN;

***check autocorrelation;
PROC AUTOREG data=ITS_spine;
model LOS_spine=month / method=ml  dwprob;
output out=out p=Prediction r=residual;
RUN;

**check for seasonality;
PROC ARIMA data=ITS_spine;
  identify var=LOS_spine stationarity=(dickey=0);
 quit; run;
```

```
**adjust for auto correlation;
PROC AUTOREG data=ITS_spine;
model LOS_spine=month hospital month_after / method=ml nlag=12 backstep
dwprob loglikl;  *adjust for age sex race payer risk soi source;
output out=out1 pm=Spine r=residual;
RUN;
*draw plot;
PROC SGPLOT data=out1 NOAUTOLEGEND;
scatter x=month y=LOS_spine/markerattrs=(symbol=circlefilled size=6
color=blue);
*series x=month y=LOS/ lineattrs=(color=blue);
series x=month y=Spine / lineattrs=(color=black pattern=dash
thickness=2);
refline 24/axis=x label="Intervention" LINEATTRS=(thickness=18
color=gwh);*transparency=0.7 ;
refline 24/axis=x LINEATTRS=(color=red thickness=2 pattern=dash );
xaxis values=(0 to 48 by 1);
yaxis values=(2 to 8 by 1);
inset 'General EUH' / position = topleft border;
inset 'Specialty EUOSC' / position = topright border ;
run;


/**********segmented regressioon of COST***********/

**adjust for auto correlation;
PROC AUTOREG data=ITS_spine;
model COST_spine=month hospital month_after/ method=ml nlag=12 backstep
dwprob loglikl;
output out=out2 pm=Spine r=residual;
RUN;
*draw plot;
PROC SGPLOT data=out2 NOAUTOLEGEND;
scatter x=month y=COST_spine/markerattrs=(symbol=circlefilled size=6
color=blue);
*series x=month y=COST_spine/ lineattrs=(color=blue);
series x=month y=Spine / lineattrs=(color=black pattern=dash
thickness=2);
refline 24/axis=x  label="Intervention" LINEATTRS=(thickness=18
color=gwh);
refline 24/axis=x LINEATTRS=(color=red thickness=2 pattern=dash );
xaxis values=(0 to 48 by 1);
yaxis values=(4000 to 9000 by 1000);
inset 'General EUH' / position = topleft border;
inset 'Specialty EUOSC' / position = topright border ;
run;



************************************HIP
SURGERY************************************;

/*****create new dataset for hip surgery*****/
data thesis_hip;
 set hip;
*convert age from month to years;
```

```
 age=age/12;
*aggregate into month;
month=intck('month','21sep2006'd,AdmissionDate);   * by 2 month--
month2,by 3 month--qtr;
if month=0 then delete; *we start our study preroid from Oct 1, 2006;
if month=24 then delete; *treat month 24 as transition period.
*define two types of hospital;
if month<24 then hospital=0; else hospital=1;
*create variable for segmented regression;
if month <24 then month_after=0; else month_after=month-24;

*recode variables;
adm_source=AdmissionSourceCode+0;
if adm_source=1 then source=1;
   else if adm_source in (2,4,5,6) then source=2;
   else source=3;
ID=compress(PatientId,'','kd');
pri_payer=floor(PrimaryPayerCode/100);
if pri_payer=1 then payer=1;
  else if pri_payer in (2,3) then payer=2;
  else payer=3;
*combine groups and make race three groups;
if racecode in (3,6) then race=3; else race=racecode;
*combine groups and make ROM three groups;
risk_mortality=ROM+0;*convert char variable to num;
if risk_mortality in (3,4) then risk=3; else risk=risk_mortality;
*combine groups and make SOI three groups;
if SOI=4 then SOI=3;
*calculate cost per hospital day;
cost=totalcost/LOS;

format sexcode sex. race race. payer payer. risk rom.  soi soi.
hospital hospital. source source. ;
label age="Age (years)";
label sexcode="Sex";
label race="Race";
label payer="Primary payer";
label risk="Risk of Mortality";
label SOI="Severity of Illness";
label source="Admission Source";
run;



/*****compare demographic variables between two types of hospital*****/
*t test for age which is normal distributed;
proc TTEST data=thesis_hip sides=2 alpha=0.05 h0=0;
 class hospital ; var age LOS cost; run;
*rank sum test for LOS and cost which are skwed distributed;
proc means data=thesis_hip mean median p25 p75;
class hospital; var LOS cost; run;
proc NPAR1WAY data=thesis_hip wilcoxon;
class hospital; var LOS cost; run;
*chisquare test for categorical variables;
PROC FREQ data=thesis_hip;
tables sexcode*hospital race*hospital payer*hospital risk*hospital
soi*hospital source*hospital/chisq;
```

```
run;

*caculate how many observations are aggregated into each month;
PROC FREQ data=thesis_hip;
tables month;
run;

*first look the scatter plot of outcome vs. day---no extrem outliners;
PROC GPLOT DATA=thesis_hip;
     PLOT LOS*month=hospital ;
RUN;

*caculate average LOS by month;
PROC MEANS data=thesis_hip;
class month; var los;
run;*

/*********Time Series************/
*caculate the average LOS and Cost by month and creat dataset called
average;
PROC SQL;
create table average_hip as
 select month as month, month_after as month_after, avg(LOS) as
LOS_hip, hospital as hospital, avg(totalcost/LOS) as COST_hip,
        age as age, sexcode as sex, payer as payer, risk as risk, soi
as soi, source as source
  from thesis_hip
   group by month;
quit;
*selct the time series points and output as csv file;
PROC SORT data=average_hip; by month; run;
data ITS_hip; set average_hip; by month;
if first.month then output ITS_hip;
*if month=24 then delete;
label LOS_hip="Length of Stay (LOS)";
label COST_hip="Cost";
run;

*export the time series data to csv;
PROC EXPORT data=ITS_hip
    outfile='H:\Childrens care\thesis\time_series_hip.csv'
    dbms=csv
    replace;
run;


/*********segmented regressioon of LOS***********/
*scatter plot of aggregate time series outcome (48ponits) vs. month;
PROC GPLOT DATA=ITS_hip;
     PLOT LOS_hip*month=hospital COST_hip*month=hospital;
RUN;

**adjust for auto correlation;
PROC AUTOREG data=ITS_hip;
model LOS_hip=month hospital month_after/ method=ml nlag=12 backstep
dwprob loglikl;
output out=out3 pm=Hip r=residual;
```

```
RUN;
*draw plot;
PROC SGPLOT data=out3 NOAUTOLEGEND;
scatter x=month y=LOS_hip/markerattrs=(symbol=circlefilled size=6
color=blue);
*series x=month y=LOS/ lineattrs=(color=blue);
series x=month y=Hip / lineattrs=(color=black pattern=dash
thickness=2);
refline 24/axis=x label="Intervention" LINEATTRS=(thickness=18
color=gwh);*transparency=0.7 ;
refline 24/axis=x LINEATTRS=(color=red thickness=2 pattern=dash );
xaxis values=(0 to 48 by 1);
yaxis values=(2 to 7 by 1);
inset 'General EUH' / position = topleft border;
inset 'Specialty EUOSC' / position = topright border ;
run;



***LOS:merge spine and hip plot into one***;
data out_LOS; merge out1 out3;by month;run;

PROC SGPLOT data=out_LOS  ;
scatter x=month y=LOS_spine /markerattrs=(symbol=circlefilled size=6
color=red) legendlabel=" " ;
series x=month y=LOS_spine/ lineattrs=(color=red) legendlabel="Spine" ;
series x=month y=Spine / lineattrs=(color=black pattern=dash
thickness=2) legendlabel="Regression";

scatter x=month y=LOS_hip/markerattrs=(symbol=circlefilled size=6
color=blue) legendlabel=" " ;
series x=month y=LOS_hip/ lineattrs=(color=blue) legendlabel="Hip" ;
series x=month y=Hip / lineattrs=(color=black pattern=dash thickness=2)
legendlabel="Regression";

refline 24/axis=x label="Intervention" LINEATTRS=(thickness=18
color=lightgray);*transparency=0.7 ;
xaxis values=(0 to 48 by 1);
yaxis values=(2 to 7 by 1);
inset 'General EUH' / position = topleft border;
inset 'Specialty EUOSC' / position = topright border ;
run;



/**********segmented regressioon of COST************/
*scatter plot of aggregate time series outcome (48ponits) vs. month;
PROC GPLOT DATA=ITS_hip;
     PLOT COST*month=hospital cost_avg*month=hospital;
RUN;

**adjust for auto correlation;
PROC AUTOREG data=ITS_hip;
model COST_hip=month hospital month_after / method=ml nlag=12 backstep
dwprob loglikl;
output out=out4 pm=hip r=residual;
RUN;
```

```
*draw plot;
PROC SGPLOT data=out4 NOAUTOLEGEND;
scatter x=month y=COST_hip/markerattrs=(symbol=circlefilled size=6
color=blue);
*series x=month y=COST_hip/ lineattrs=(color=blue);
series x=month y=hip / lineattrs=(color=black pattern=dash
thickness=2);
refline 24/axis=x  label="Intervention" LINEATTRS=(thickness=18
color=gwh);
refline 24/axis=x LINEATTRS=(color=red thickness=2 pattern=dash );
xaxis values=(0 to 48 by 1);
yaxis values=(2000 to 8000 by 1000);
inset 'General EUH' / position = topleft border;
inset 'Specialty EUOSC' / position = topright border ;
run;



***COST:merge spine and hip plot into one***;
data out_COST; merge out2 out4;by month;run;
PROC SGPLOT data=out_COST;
scatter x=month y=COST_spine/markerattrs=(symbol=circlefilled size=6
color=red) legendlabel=" " ;
series x=month y=COST_spine/ lineattrs=(color=red) legendlabel="Spine"
;
series x=month y=spine / lineattrs=(color=black pattern=dash
thickness=2)legendlabel="Regression";

scatter x=month y=COST_hip/markerattrs=(symbol=circlefilled size=6
color=blue) legendlabel=" ";
series x=month y=COST_hip/ lineattrs=(color=blue) legendlabel="Hip" ;
series x=month y=hip / lineattrs=(color=black pattern=dash thickness=2)
legendlabel="Regression";

refline 24/axis=x  label="Intervention" LINEATTRS=(thickness=18
color=gwh);
xaxis values=(0 to 48 by 1);
yaxis values=(2000 to 9000 by 1000);
inset 'General EUH' / position = topleft border;
inset 'Specialty EUOSC' / position = topright border ;
run;




************************************aggregate by 2
months******************************;
data thesis_hip;
 set hip;
*convert age from month to years;
 age=age/12;
*aggregate into month;
month=intck('month2','21sep2006'd,AdmissionDate);   * by 2 month--
month2,by 3 month--qtr;
*if month=0 then delete; *we start our study preroid from Oct 1, 2006;
*if month= then delete;
*define two types of hospital;
```

```
if month<12 then hospital=0; else hospital=1;
*create variable for segmented regression;
if month <12 then month_after=0; else month_after=month-11;

*recode variables;
adm_source=AdmissionSourceCode+0;
if adm_source=1 then source=1;
   else if adm_source in (2,4,5,6) then source=2;
   else source=3;
ID=compress(PatientId,'','kd');
pri_payer=floor(PrimaryPayerCode/100);
if pri_payer=1 then payer=1;
  else if pri_payer in (2,3) then payer=2;
  else payer=3;
*combine groups and make race three groups;
if racecode in (3,6) then race=3; else race=racecode;
*combine groups and make ROM three groups;
risk_mortality=ROM+0;*convert char variable to num;
if risk_mortality in (3,4) then risk=3; else risk=risk_mortality;
*combine groups and make SOI three groups;
if SOI=4 then SOI=3;

format sexcode sex. race race. payer payer. risk rom.  soi soi.
hospital hospital. source source. ;
label age="Age (years)";
label sexcode="Sex";
label race="Race";
label payer="Primary payer";
label risk="Risk of Mortality";
label SOI="Severity of Illness";
label source="Admission Source";
run;


*caculate average LOS by month;
PROC MEANS data=thesis_hip;
class month; var los;
run;*

*caculate the average LOS and Cost by month and creat dataset called
average;
PROC SQL;
create table average_hip as
 select month as month, month_after as month_after, avg(LOS) as LOS,
hospital as hospital, avg(totalcost/LOS) as cost,
        age as age, sexcode as sex, payer as payer, risk as risk, soi
as soi, source as source
  from thesis_hip
   group by month;
quit;
*selct the time series points and output as csv file;
PROC SORT data=average_hip; by month; run;
data ITS_hip; set average_hip; by month;
if first.month then output ITS_hip;
*if month=24 then delete;
label LOS="Length of Stay (LOS)";
label COST="Cost";
```

```
run;

*segmented regressioon of LOS;
*scatter plot of aggregate time series outcome (24ponits) vs. month;
PROC GPLOT DATA=ITS_hip;
     PLOT LOS*month=hospital cost*month=hospital;
RUN;

**adjust for auto correlation;
PROC AUTOREG data=ITS_hip;
model LOS=month hospital month_after/ method=ml nlag=12 backstep dwprob
loglikl;
output out=out pm=prediction r=residual;
RUN;
*draw plot;
PROC SGPLOT data=out NOAUTOLEGEND;
scatter x=month y=LOS/markerattrs=(symbol=circlefilled size=6
color=blue);
series x=month y=LOS/ lineattrs=(color=blue);
series x=month y=prediction / lineattrs=(color=black pattern=dash
thickness=2);
refline 12/axis=x label="Intervention" LINEATTRS=(thickness=18
color=gwh);*transparency=0.7 ;
xaxis values=(0 to 24 by 1);
yaxis values=(2 to 7 by 1);
inset 'General EUH' / position = topleft border;
inset 'Specialty EUOSC' / position = topright border ;
run;

*segmented regressioon of COST;
*scatter plot of aggregate time series outcome (24ponits) vs. month;
PROC GPLOT DATA=ITS_hip;
     PLOT COST*month=hospital cost*month=hospital;
RUN;

**adjust for auto correlation;
PROC AUTOREG data=ITS_hip;
model COST=month hospital month_after/ method=ml nlag=12 backstep
dwprob loglikl;
output out=out pm=prediction r=residual;
RUN;
*draw plot;
PROC SGPLOT data=out NOAUTOLEGEND;
scatter x=month y=COST/markerattrs=(symbol=circlefilled size=6
color=blue);
series x=month y=COST/ lineattrs=(color=blue);
series x=month y=prediction / lineattrs=(color=black pattern=dash
thickness=2);
refline 12/axis=x  label="Intervention" LINEATTRS=(thickness=18
color=gwh);
xaxis values=(0 to 24 by 1);
yaxis values=(2000 to 8000 by 1000);
inset 'General EUH' / position = topleft border;
inset 'Specialty EUOSC' / position = topright border ;
run;
```