Proceedings of the National Conference
On Undergraduate Research (NCUR) 2015
Eastern Washington University, Cheney, WA
April 16-18, 2015

A Discussion about the Healthcare Costs and Insurance Purchase

Xinhe (Tina) He
Operations Management
Fisher College of Business, The Ohio State University
820 Fisher Hall, 2100 Neil Avenue
Columbus, OH 43210

Advised By: Dr. Kewei Hou

Abstract

The U.S. Healthcare system is one of the most expensive healthcare system in the world. The health care sector takes 16.9% of U.S. GDP, while this ratio is 9.3% on average in other countries (Source: Organization for Economic Co-operation and Development). The nation's health care expenditure reached $2.8 trillion in 2012, or $8,915 per person (Source: center for Medicare and Medicaid Services). To solve the healthcare affordability issue, the U.S. government has been continuously making efforts to make insurance more affordable. In 2009, Affordable Care Act (ACA) was launched and its main action was to have a universal insurance plan. However, it is an understudied area about how this will help reduce healthcare costs. This research attempts to find statistical evidence that healthcare costs is related to insurance coverage increase. In this research, multiple regression tests, T-tests, and robustness analyses are used to analyze the pre-ACA insurance market in 1999-2009. This study is able to statistically confirm a positively correlating relationship exists between healthcare cost and insurance coverage, and this suggests that government intervention is necessary in the insurance market, because the market is less likely to reach a demand and supply balance point.

Keywords: healthcare, insurance coverage, healthcare costs, state variation, year variation, consumer psychology, public health

1. Introduction

The U.S. government has been continuously making health reforms to benefit more public members - just name a few Acts: Health Insurance Portability and Accountability Act (HIPAA) 1996, Health Center Growth Initiative 2002, Deficit Reduction Act 2005. However, despite those efforts, but the U.S. healthcare costs were surging at a rapid rate since 1990s (see Figure 1). One example is the insurance premium cost, the increase rate is at least 50 percent from 2003 to 2010 for family coverage. If it is maintained at this rate, by 2020 the average premium for family coverage will reach $24,000 (Source: Realizing Health Reform's Potential, 2011).

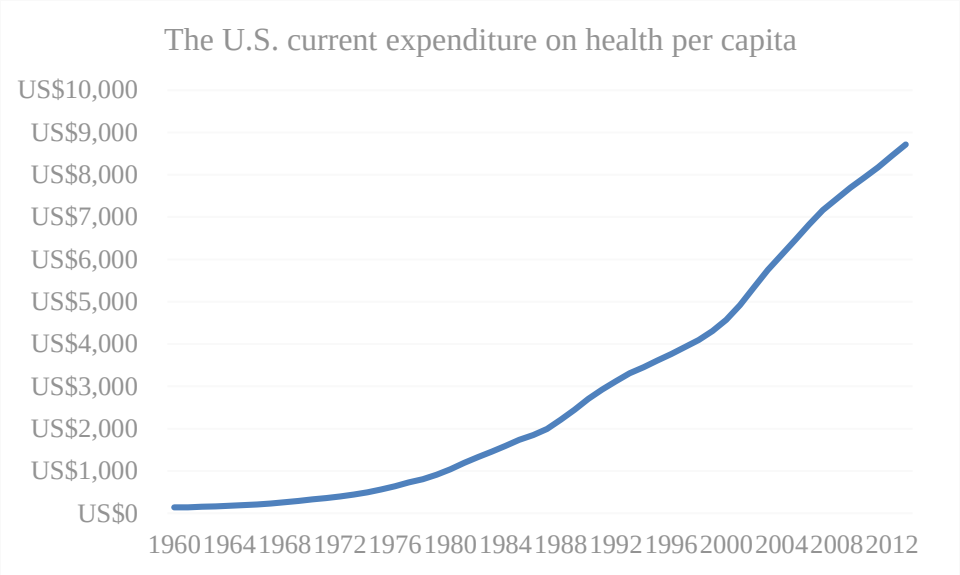

Figure 1. U.S. healthcare costs are surging rapidly after 1990s

Aware of the cost issue, the U.S. government launched ACA in 2009 and attempted to control insurance and healthcare costs by enacting new legal measures. One of ACA measures was to have universal insurance coverage. While it is still too new to conclude how much contribution this act is making, this research explores whether universal insurance coverage will help bring down healthcare costs. The study examines the relationship between healthcare cost and insurance coverage under a pre-ACA insurance market from 1991-2009 in 51 regions. It attempts a statistical approach to answer the following question: do higher health care costs encourage or discourage consumers' health insurance purchases?

2. Methodology

The research studies healthcare variations from 1991-2009 before ACA was launched. Key variables are calculated through demographic data such as healthcare insurance coverage, average health spending, and the average individual income for 51 U.S. regions (50 states and Washington D.C.). The data were sourced from various creditable public healthcare sources and were organized into a relational database for analysis purposes. For the analysis, it takes three steps: multiple regression, one-sample T test, and robustness analysis.

The multiple regression analysis aims to find the correlating relationship when one of the external variables (i.e. *Year* or *State*) remains constant. It assumes that *State* and *Year* factors impact each other. The regression analysis attempts to explain the insurance coverage variation by the healthcare cost variable along with other independent variables, including poverty level, medicare insurance percentage, and state average age. The within-state analysis explained that how variation of healthcare costs can explain coverage variation for a particular state, while the across-state analysis looks at variations at all states in one year. To be more specific, in the within-state analysis, the *State* variable remains constant and each state is analyzed individually with 19 years insurance coverage data. Oppositely, in the across-state analysis, the *Year* variable remains constant and each year is analyzed individually with 51 regions insurance coverage data. In both within-state and across-state analyses, a positive coefficient indicates that insurance purchase demand increases with rising healthcare costs; a negative coefficient indicates that insurance purchase demand decreases with rising healthcare costs.

The T-test analysis is based on the regression correlation outputs. It statistically tests whether a significant relationship exists between the health and cost variables. To confirm this relationship, the T-test hypothesizes the coefficient average equals to 0 and generates p value to reject the hypothesis. When p value is less than the alpha value (0.05 at the initial stage of the research and 0.01 after removing the outliers), the T-test is able to reject the hypothesis successfully and hence to confirm the relationship between variables.

The robustness analysis is used to further refine the analysis by eliminating outliers from the regression test, rerunning the T-test, and providing more statistical significance even at under a higher confidence interval requirement

(at 99% Confidence Interval).

2.1 Research Design

This research hypothesizes that insurance coverage is related to healthcare costs and the data analysis is based on the following variables:

Dependent variable: insurance coverage percentage (i.e. insured population ÷ All population, by state),
Independent variable: healthcare cost percentage (i.e. Average healthcare costs ÷ Average income, by state), poverty percentage (percentage of population living below poverty line), state average age, and medicare coverage percentage.

The relevancy between the variables is examined by finding correlation significance through both within-state and across-state approaches. If either of these two approaches is successful, we can confirm the hypothesis and conclude that insurance coverage and healthcare costs are correlated.

3. Test Results

3.1 Multiple Regression Analysis Model Results

3.1.1 multiple regression analysis is made up of two components: within state analysis and across state analysis.

3.1.2 within state

State analysis is made up of 51 data points represented by 51 states' regression model coefficient. In each regression model, 19 data points from 1991-2009 are included in each state regression model to see state variation. As shown in Figure 2, distinctive difference of coefficient values are observed in the within-state analysis. The coefficient average is 0.1406, which is a positive value that indicates that people tend to buy more insurance when healthcare costs are surging.

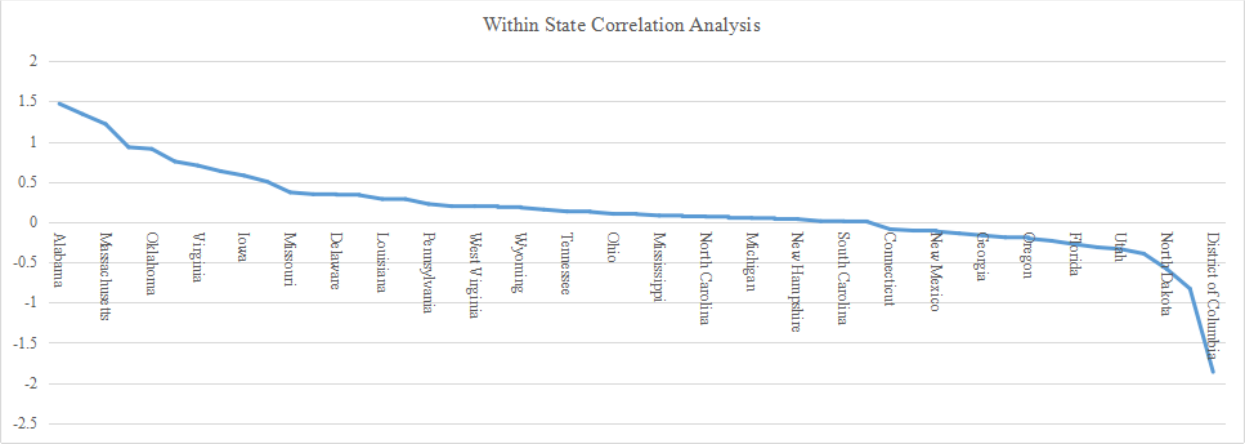

Figure 2. Across State Correlation Analysis

3.1.3 across state analysis

The across state analysis is made up of 19 data points represented by 19 years' regression model coefficient. In each regression model, 51 data points from 51 states are regressed to see the same-year regression relevance across state.

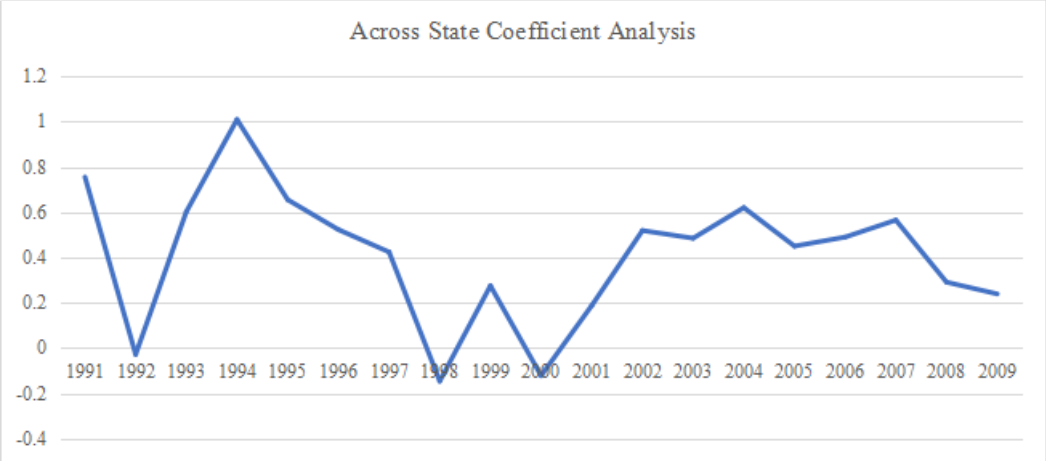

Figure 3. Across State Correlation Analysis

As shown in Figure 3, the coefficient value is observed in the across-state analysis and the average is 0.4102 in linear regression models.

4. T-test Results

In the T-test, the hypothesis is that the coefficient average equals to 0. The p value from T-tests can reject the hypothesis if it is smaller than alpha value, and this analysis shows significance relationship between the variables; otherwise, if the p value is larger than or equal to the alpha value, there is no statistical evidence to draw conclusions. As shown in figure 4, under 95% confidence interval (alpha value = 0.05), the within-state analysis T-test result, but not the across-state analysis, shows statistical significance (see Figure 4).

One-Sample T: Multiple regression coefficient_Across State Analysis
Test of $\mu = 0$ vs $\neq 0$

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
Year coefficient	19	0.4102	0.2959	0.0679	(0.2676, 0.5528)	6.04	0

One-Sample T: Multiple regression coefficient_Within State Analysis
Test of $\mu = 0$ vs $\neq 0$

Variable	N	Mean	StDev	SE Mean	95% CI	T	P
State coefficient	51	0.1406	0.5341	0.0748	(-0.0097, 0.2908)	1.88	0.066

Figure 4. T-test results (95% confidence interval)

As shown in figure 4, in the across-state analysis, significant relationships can be found between the insurance coverage variable and healthcare costs variable (p value= 0< 0.05) under 95% Confidence Interval; however, the within-state analysis(p value= 0.066>0.05) does not show statistical significance under the same confidence level. The T test result difference between these two analyese shows that study method can be a key consideration for this study.

4.1 Robustness Analysis Results

The robustness analysis removes outliers, refines research accuracy, and focuses on the large majority of data that represents the common reality. This research identifies the D.C. district correlation as an unusual observation given its distinctive negative coefficient value from the rest of the data points (see Figure 2). When the outlier is ignored, the T-test can increase the confidence interval for both methods as shown in figure 5.

One-Sample T: Multiple regression coefficient_Across State Analysis_without outlier (99% CI) Test of μ = 0 vs ≠ 0							
Variable	N	Mean	StDev	SEMean	99% CI	T	P
Year coefficient without outlier	19	0.5248	0.2941	0.0675	(0.3306, 0.7191)	7.78	0.000

One-Sample T: Multiple regression coefficient_Within State Analysis_without outlier (99% CI) Test of μ = 0 vs ≠ 0							
Variable	N	Mean	StDev	SEMean	99% CI	T	P
State coefficient without outlier	50	0.1806	0.4557	0.0644	(0.0079, 0.3533)	2.8	0.007

Figure 5. The modified T-test results further improved statistical test significance

As shown in figure 5, after removing outliers for both analyses, evidence shows that healthcare cost variable is correlated with insurance coverage under 99% confidence interval (p value <0.01). Therefore, the hypothesis is confirmed - we are 99% confident that the healthcare costs are related to insurance coverage. Both coefficient averages are positive, suggesting insurance purchase behaviors are encouraged when healthcare costs increase. This makes sense from a healthcare consumer's standpoint, because when healthcare costs are getting higher, people tend to buy insurance to get more coverage, so that insurance companies can share their financial burden of the healthcare expenses.

4.2 Research Insights

This research contributes to future research topics such as statistical study method and healthcare cost prediction multiple regression model.

4.2.1 study method

This research contributes on how to measure and compare variables' influence. When a factor is too conceptual to be measured, it can be designed as a fixed control variable to examine the test results. This would help structure analysis when a specific variable is influential but complicated. For example, in the within-state analysis, the state variable is set unchanged, so that state situation variation will be excluded to avoid complication; on the contrary, state variation is included for across-state analysis for research comprehensiveness.

4.2.2 insurance market analysis

Insurance coverage tends to increase over the years and the purchase might be made through either public or private funding. If more coverage comes from public insurance coverage, it means more public funds are used in getting insurance coverage instead of in economic development; if these increased coverage comes from private insurance, it creates financial and psychological stress on healthcare consumers, and anxious customers would drive higher demands, even higher premium prices, and hence higher healthcare costs as a result. In either case, continuously rising healthcare cost is not desirable.

The correlation analysis suggests that the insurance market is a flawed demand-driven market - demand and supply for insurance will not balance automatically to reach a market equilibrium point. If government intervention is not present, the insurance market should not be able to contain the cost surge by itself. s

4.2.3 quantified relationship

This research is meaningful in quantifying the relationship between the insurance coverage and healthcare costs. For instance, in within-state analysis, when cost percentage increases by 1%, the insurance coverage percentage is expected to increase by about 0.5%, or 50 basis points (e.g. from 18% to 18.5%). This quantified relationship will be able to help future researchers to relate the insurance coverage with other control variables for to relate the coverage ratio to other control variables. Similarly, studies can quantify the correlation of healthcare coverage and other variables. These will provide valuable information to understand how much influence each variable has on insurance coverage variation, so that funding can be distributed strategically to improve healthcare affordability situation.

4.2.4 potential model improvement

Potential model improvements can be made to increase the research comprehensiveness. One improvement is to regress the previous year insurance cost percentage and the next year coverage percentage, e.g. 1991 healthcare cost as the independent variable & 1992 insurance coverage as the dependent variable. Other improvements can be made to include more sophisticated variables to further explain the insurance coverage variation over the years, e.g. percentage of private insurance, insurance policy variation across states, etc.

5. Conclusion

This research finds that the insurance coverage is positively correlated with healthcare cost variable. From the study, we can be 99% confident that insurance coverage is correlated with healthcare costs, which suggests that the insurance purchase demand is encouraged by rising healthcare costs. In other words, in health insurance market, market equilibrium will hardly be achieved by balancing purchase demand and supply. One benefit of ACA is that it regulates the insurance market demand, and hence prevents demand-driven healthcare cost surge from insurance premium price increase.

Besides confirming that healthcare cost variable and insurance coverage variable are correlated, this research is also unique in justifying the U.S. insurance market with different methods (Within States and Across States methods), quantifying variables influences, and potentially contributing to predictive analytics model development. These will be meaningful to the healthcare affordability improvement progress and healthcare public policy making.

6. Reference

1. Kaiser family foundation, “health costs by states database”, http://kff.org/statedata/
2. Integrated Public Use Microdata Series (IPUMS), “Current Population Survey (CPS) insurance survey database”, https://cps.ipums.org/cps/
3. U.S. Dept. of Commerce, Bureau of Economic Analysis (2013 release) and prepared by Bureau of Business & Economic Research, university of New Mexico - per capital personal income by state https://bber.unm.edu/econ/us-pci.htm
4. United States Census Bureau, data ferret analysis tool http://dataferrett.census.gov/
5. Organization for Economic Co-operation and Development, “National Health Statistics 2014”, http://www.oecd.org/els/health-systems/health-data.htm
6. Kaiser family foundation, “history of health reform”, https://kaiserfamilyfoundation.files.wordpress.com/2011/03/5-02-13-history-of-health-reform.pdf
7. U.S. Healthcare expenditure GDP percentage, Organization for Economic Co-operation and Development - National Health Statistics 2014, accessed on January 10, 2015
8. 2012 health spending per person, Center for Medicare and Medicaid Services, accessed on January 10, 2015
9. The U.S. current expenditure on health per capita, Organisation for Economic Co-operation and Development 2015 (Accessed on November 10, 2015)