

2007

# Organizational form of disease management programs: A transaction cost analysis

Nahush Chandaver  
*University of South Florida*

Follow this and additional works at: <http://scholarcommons.usf.edu/etd>

 Part of the [American Studies Commons](#)

---

## Scholar Commons Citation

Chandaver, Nahush, "Organizational form of disease management programs: A transaction cost analysis" (2007). *Graduate Theses and Dissertations*.  
<http://scholarcommons.usf.edu/etd/665>

This Thesis is brought to you for free and open access by the Graduate School at Scholar Commons. It has been accepted for inclusion in Graduate Theses and Dissertations by an authorized administrator of Scholar Commons. For more information, please contact [scholarcommons@usf.edu](mailto:scholarcommons@usf.edu).

Organizational Form of Disease Management Programs: A Transaction Cost Analysis

by

Nahush Chandaver

A thesis submitted in partial fulfillment  
of the requirements for the degree of  
Master of Science in Industrial Engineering  
Department of Industrial and Management Systems Engineering  
College of Engineering  
University of South Florida

Major Professor: Kingsley A. Reeves, Jr., Ph.D.  
Michael Weng, Ph.D.  
Gabriel Picone, Ph.D.

Date of Approval:  
November 14, 2007

Keywords: predictive modelling, logistic regression, selection bias, inverse mills ratio,  
outsourcing, integration

© Copyright 2008, Nahush Chandaver

## Table of Contents

List of Tables .....	iv
List of Figures .....	vii
Abstract .....	viii
Chapter 1. Introduction .....	1
1.1 Introduction to the Disease Management Concept .....	1
1.2 Outsourcing in the Health Insurance Industry .....	4
1.3 Background, Complication and Objectives .....	4
1.4 Research Approach and Benefits .....	5
1.5 Structure of the Report.....	8
Chapter 2. Literature Review .....	9
2.1 Introduction.....	9
2.2 Transaction Cost Economics (TCE) Theory .....	9
2.2.1 Uncertainty.....	11
2.2.2 Asset Specificity and its Effect on Organizational Form.....	15
2.2.3 Similarity and Frequency .....	18
2.2.4 Bounded Rationality .....	20
2.2.5 Core Competence and Transaction Costs .....	21
2.2.6 Empirical Measurement of Transaction Costs.....	24
2.3 Disease Management Literature Review .....	25
2.3.1 A Brief History of Disease Management.....	26
2.3.2 Current State of the Disease Management Industry.....	27
2.3.3 Effect of Transaction Cost Factors on DM Organizational Form.....	34
2.3.4 Future of Disease Management .....	40
2.4 Selection Bias and the Heckman Two-Step Method .....	41
Chapter 3. Literature Summary and the Hypotheses .....	44
3.1 Literature Summary and Application to Health Plans .....	44
3.2 Application of TCE Factors to Health Plans.....	45
3.3 The Hypotheses.....	48
Chapter 4. Methodology .....	49
4.1 Research Approach .....	49
4.2 Research Method .....	49
4.3 Design and Study Participants .....	52
4.4 Measures, Instruments and Data Sources .....	53
4.5 Primary Data Collection .....	54
4.6 Data Analysis .....	57

Chapter 5. Numerical Results and Inference .....	59
5.1 Frequencies of Respondents and Corresponding DM Programs .....	59
5.2 Data Analysis .....	64
5.3 Creating the Training and Validation Sets for the First Stage Selection Model .....	67
5.4 Training Stage.....	68
5.4.1 Step 1. Creating the First Stage Selection Model .....	68
5.4.1 Step 2. Evaluate Results of the Training Stage.....	73
5.4.1 Step 3. Use the Classification Table to Determine Optimal Cut-Off Point .....	77
5.5 Validation Stage.....	78
5.5.1 Step 1. Embed the Validation Set Into the Training Set .....	78
5.5.1 Step 2. Results of the First Stage Selection Model With Combined Data Set .....	82
5.5.2 Classification of the Training Set .....	83
5.5.3 Classification of the Validation Set .....	84
5.5.4 Inference for the First Stage Selection Model .....	84
5.6 First Stage Selection Model Excluding Uncertainty .....	86
5.6.1 Training Stage.....	87
5.6.1 Step 1. Creating the First Stage Selection Model Excluding Uncertainty ..	87
5.6.1 Step 2. Evaluate Results of the Training Stage for the Seven Factor Model .....	89
5.6.1 Step 3. Use the Classification Table to Determine Optimal Cut-Off Point .....	91
5.6.2 Testing Stage.....	92
5.6.2 Step 1. Embed the Validation Set Into the Training Set .....	92
5.6.2 Step 2. Results of the First Stage Selection Model With Combined Data Set for Seven TCE Factors .....	95
5.6.3 Classification of the Training Set .....	96
5.6.4 Classification of the Validation Set .....	97
5.7 Comparison of the Model With and Without the TCE Factor Uncertainty .....	97
5.8 Inference From the Comparison of Full and Seven Factor Model.....	98
5.9 Calculation of Actual In-Sourced and Outsourced Costs .....	99
5.9.1 Breakup of Actual In-Sourced Costs .....	99
5.9.2 Breakup of Actual Outsourced Costs.....	100
5.9.3 Calculation of the Inverse Mills Ratio and the Help and Control Factor Delta.....	101
5.10 Frequencies for the In-Sourced Subset .....	104
5.10.1 Frequencies for the Organizations in the In-Sourced Subset.....	104
5.10.2 Frequencies for the Diseases in the In-Sourced Subset .....	105
5.11 Organization Cost Model for In-Sourced Costs.....	106
5.11.1 Running the Model .....	107
5.11.2 Results of the In-Sourced Cost Model With Correction for Selection Bias .....	109
5.11.3 Correcting the Standard Error Terms.....	111
5.11.4 Inference From the Internal Cost Equation.....	116
5.12 Comparison of First and Second Stage Results .....	118
5.13 Comparison of Actual and Predicted In-Sourced Costs.....	120
5.13.1 Means for the Actual and Predicted In-Sourced Costs .....	120
5.13.2 Comparison of Actual and Predicted Costs for Sub-Sample of Data.....	121

5.13.3 Rolling Up the Costs for Each Organization in the In-Sourced Subset.....	121
5.14 Creating the Log Specification Model for the In-Sourced Costs.....	122
5.14.1 Running the Model .....	122
5.14.2 Results of the Log Specification In-Sourced Cost Model.....	124
5.14.3 Correcting the Standard Errors for the Log Spec Model .....	125
5.15 Comparison of Actual and Predicted In-Sourced Costs.....	129
5.15.1 Means for the Actual and Predicted In-Sourced Costs From the Log Specification Model .....	130
5.15.2 Comparison of Actual and Predicted Costs for Sub-Sample of Data From the Log Specification Model .....	131
5.15.3 Combined Actual Cost, Combined Predicted Cost and Combined Error Estimate Using the Log Specification of the Cost Model .....	132
5.15.4 Cost and Error Estimates for the Whole In-Sourced Subset Using the Log Specification of the Cost Model .....	133
5.15.5 Comparison of Coefficients From the Regular and Log Specification of the In-Sourced Cost Model.....	135
5.15.6 Comparison of Summed Actual and Predicted Costs From the Regular and Log Specification of the In-Sourced Cost Model.....	136
Chapter 6. Conclusions and Future Research .....	138
6.1 Conclusions.....	138
6.2 Organizational Form and Cost Analysis and Prediction as a Decision Making Tool .....	139
6.3 Future Work.....	140
References.....	142
Appendices.....	149
Appendix A. SAS Code .....	150
Appendix B. Electronic Survey .....	170

## List of Tables

Table 2.1 Disease Management Program Statistics Across the U.S. ....	33
Table 2.2 Method of Disease Management Program Implementation .....	33
Table 4.1 Probit Model Variable Definitions and Descriptions.....	55
Table 5.1 Responding Organizations and Number of Respective Responses.....	60
Table 5.2 Frequency of Corresponding DM Programs.....	61
Table 5.3 Frequencies of the Organization Form .....	62
Table 5.4 Frequencies of the Eight TCE Factors .....	63
Table 5.5 Means and Statistics for all Modeling Variables .....	65
Table 5.6 Correlations for all Modeling Variables .....	66
Table 5.7 Frequencies for the Training Set.....	68
Table 5.8 Frequencies for the Validation Set.....	68
Table 5.9 Full Model Response, Fit Statistics and Null Hypothesis for Training Stage .....	70
Table 5.10 Analysis of Parameter Coefficients for the Training Stage .....	76
Table 5.11 Classification Table for Training Set.....	78
Table 5.12 Full Model Response, Fit Statistics and Null Hypothesis for Testing Phase.....	80
Table 5.13 Analysis of Parameter Coefficients for Testing Stage .....	82
Table 5.14 Prediction Accuracy for the Training Set .....	83
Table 5.15 Prediction Accuracy for the Validation Set .....	84
Table 5.16 Seven Factor Model Response, Fit Statistics and Null Hypothesis for Training Stage .....	87
Table 5.17 Analysis of Parameter Coefficients for the Seven Factor Model.....	90

Table 5.18 Classification Table for the Seven Factor Training Stage .....	92
Table 5.19 Seven Factor Model Response, Fit Statistics and Null Hypothesis for Testing Stage .....	93
Table 5.20 Analysis of Parameter Coefficients for the Testing Stage of the Seven Factor Model .....	95
Table 5.21 Prediction Accuracy for the Training Set .....	96
Table 5.22 Prediction Accuracy for the Validation Set .....	97
Table 5.23 Comparison of the Full and Seven Factor Models.....	98
Table 5.24 Means for the Actual In-Sourced Costs .....	100
Table 5.25 Means for Actual Outsourced Costs .....	101
Table 5.26 Results and Statistics for the Inverse Mills Ratio and Control Factor Delta .....	103
Table 5.27 Frequencies for Responding Organizations of the In-Sourced Subset.....	105
Table 5.28 Frequencies for the DM Programs of the In-Sourced Subset .....	106
Table 5.29 Analysis of Variance for the In-Sourced Cost Model.....	109
Table 5.30 Parameter Estimates for the Independent Variables in the In-Sourced Cost Model .....	111
Table 5.31 Values of Corrected Variance, Std. Error and Error Correlation.....	113
Table 5.32 Results From Heteroskedasticity Correction .....	114
Table 5.33 Final Cost Model Results.....	116
Table 5.34 Comparison of Selection and Substantial Model Coefficients .....	119
Table 5.35 Means and Statistics for Actual and Predicted In-Sourced Costs.....	120
Table 5.36 Comparison Between Actual and Predicted Costs for Sub-Sample of Data.....	121
Table 5.37 Rolling Up the Costs for Each Organization in the Integrated Subset.....	122
Table 5.38 Analysis of Variance for Log Specification of In-Sourced Cost Model.....	123
Table 5.39 Parameter Estimates for the In-Sourced Log Specification Model.....	125
Table 5.40 Log Specification Corrected Variance, Std. Error and Error Correlation.....	126

Table 5.41 Log Spec Heteroskedasticity Correction Results.....	127
Table 5.42 Final Log Spec Cost Model Results.....	129
Table 5.43 Means and Statistics for Actual and Predicted In-Sourced Costs From the Log Specification Model .....	131
Table 5.44 Comparison of Actual and Predicted Costs for Sub-Sample of Data From the Log Specification Model.....	132
Table 5.45 Summing Up the Actual and Predicted Costs and the Error Estimate for Each Organization.....	133
Table 5.46 Cost and Error Estimates for the Full In-Sourced Subset From the Log Specification Model .....	134
Table 5.47 Coefficient Comparison Between Standard and Log Specification of the Cost Model .....	135
Table 5.48 Comparison of Costs From the Standard and Log Specification of the Cost Model .....	136

## List of Figures

Figure 1.1 Disease Management Supply Chain .....	3
Figure 2.1 Flexibility Need v/s Control Need Option Range.....	13
Figure 2.2 Competitive Advantage v/s Strategic Vulnerability Matrix .....	14
Figure 2.3 Percentage of DM Programs Offered by Health Plans in the U.S.....	29
Figure 2.4 Organizational Form of DM Programs in Health Plans .....	35
Figure 2.5 Percentage of Health Plans and Respective Anticipated Areas of Expansion for DM Programs .....	40

## **Organizational Form of Disease Management Programs: A Transaction Cost Analysis**

Nahush Chandaver

### **ABSTRACT**

Patient care programs such as wellness, preventive care and specifically disease management programs, which target the chronically ill population, are designed to reduce healthcare costs and improve health, while promoting the efficient use of healthcare resources, and increasing productivity. The organizational form adopted by the health plan for these programs, i.e. in-sourced vs. outsourced is an important factor in the success of these programs and the extent to which the core objectives listed above are fulfilled.

Transaction cost economics aims to explain the working arrangement for an organization and to explain why sourcing decisions were made by considering alternate organizational arrangements and comparing the costs of transacting under each. This research aims to understand the nature and sources of transaction costs, how they affect the sourcing decision of disease management and other programs, and its effect on the organization, using current industry data. Predictive models are used to obtain empirical results of the influence of each factor, and also to provide cost estimates for each organizational form available, irrespective of the form currently adopted.

The analysis of the primary data obtained by the means of a web-based survey supports and confirms the effect of transaction cost factors on these programs. This implies that in order to reap financial rewards and serve patients better, health plans must aim to minimize transaction costs and select the organizational form that best accomplishes this objective.

## **Chapter 1. Introduction**

This is a thesis on the concept of outsourcing in the health insurance industry with a specific focus on disease management programs operated by the various health plans. It aims to answer the question if sourcing of disease management programs can be explained based on transaction cost factors and used to lead to cost savings.

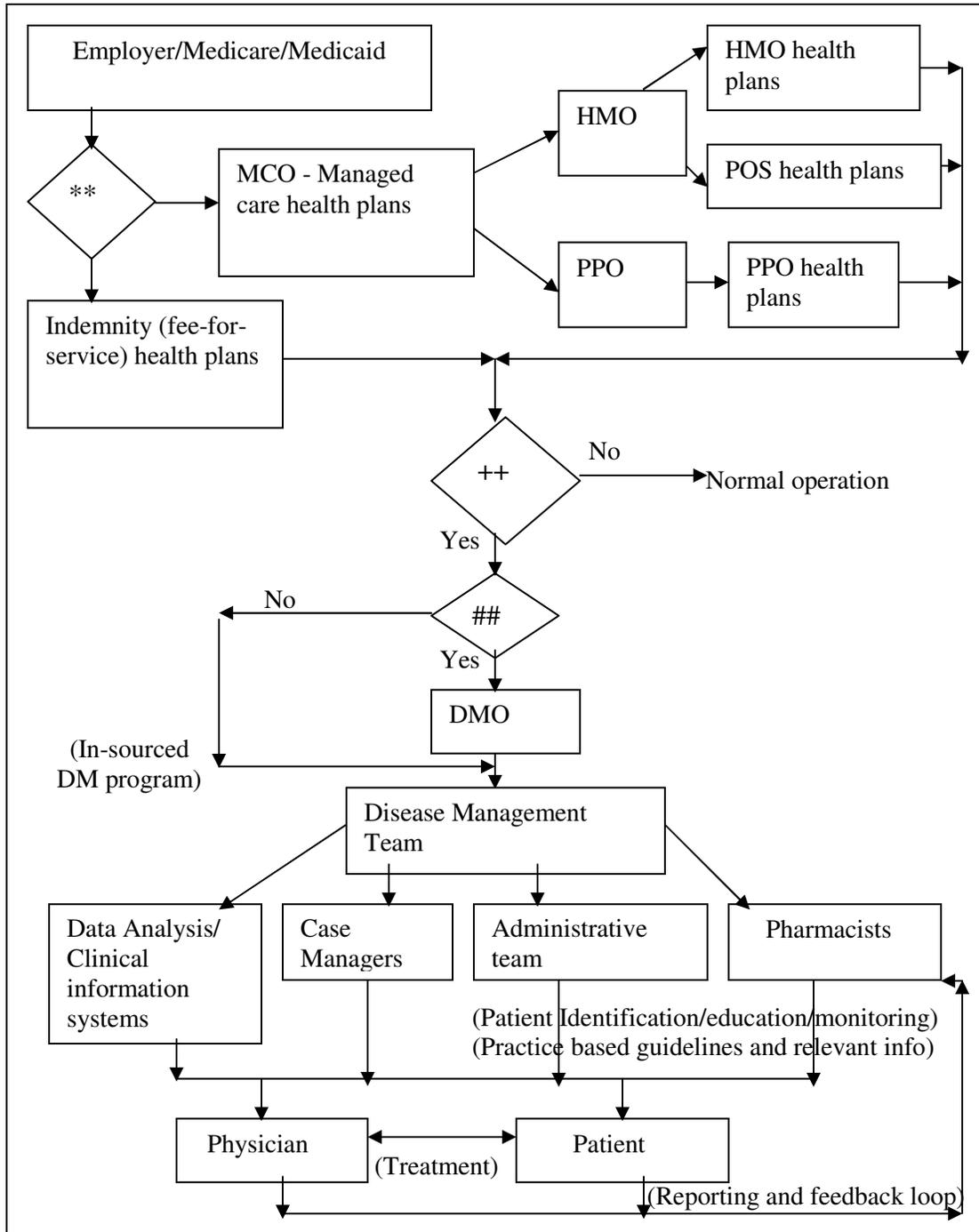
### **1.1 Introduction to the Disease Management Concept**

The quality of healthcare and health services has been the subject of public scrutiny and much debate, and it has recently heightened due to the rapid growth of costs and litigation in the form of lawsuits for negligence. There is increasing dissatisfaction of healthcare consumers with their experience due to significant deviations from best care practices, rise in medical errors and a large addition of unknown or non-value added services in healthcare [5]. A lingering concern is the inability of the U.S. healthcare system to deal with the chronically ill population, which has been increasing in recent years [39]. As noted by the Florida Medicaid Disease Management Initiative in 2000 [69], disease management programs have been proposed in order to improve healthcare by facilitating and addressing several key issues outlined below.

Disease management programs are designed to benefit both the healthcare organization and the patient by following a two-pronged approach. At the patient side, the chronically ill and the population at risk for chronic diseases are admitted in these programs. The program then takes steps to improve the health outcomes and quality of life for the patient. It does this by fostering self-care/self management of the condition by the patients themselves, aided by patient education and by raising the awareness of the patient regarding his or her own health conditions. Doing so also promotes accountability of the patient in the care and treatment decisions taken. As the awareness of the patient regarding the condition(s) is increased, it leads to a more beneficial and stronger relationship between the physician and the patient. The program staff undertakes patient monitoring and promotes the continuity of care, that is, takes steps to ensure that the patient

completes the entire treatment cycle and also measures patient satisfaction and treatment effectiveness for each patient on an ongoing basis. At the physician/care provider side, these programs aid the medical professionals by providing them valuable relevant information and practice/evidence-based guidelines that may prove helpful to them during patient treatment and care. By doing so these programs can delay and, in the best cases, even prevent complications of chronic health conditions. This leads to an improvement in the health outcomes and quality of life for the patient, while at the same time it leads to cost savings for the patient in terms of healthcare costs, and also for the healthcare provider, and is thus very beneficial to all parties involved. Disease management programs also promote efficient use of healthcare resources and increase medical productivity by increasing patient awareness levels and helping physicians in their treatment protocol. The supply chain for a disease management program is as shown in figure 1 below.

Disease management programs are particularly applicable and useful to Florida as it is the third largest in Medicaid spending in the U.S and ranks 41<sup>st</sup> in the nation in per capita expenditures [17]. The state of Florida is also a pioneer in this area as it is the first state to implement these programs in the Medicare and Medicaid fields in 1998 and encourage health plans to adopt these plans at the same time. The state government has already reduced the annual budget for Medicare and Medicaid by \$ 66 million in anticipation of the savings that were promised by the proponents of these programs, and the results of the early studies done to measure the effectiveness and results of these programs [43]. However, subsequent findings have shown that while savings in healthcare costs have occurred for patients, they have been offset to some extent by rising drug costs. Moreover, the savings for the health management organizations have been offset by the cost of implementation of the disease management programs. This has reduced the actual savings and effectiveness of the programs in terms of efficiency and cost savings for the healthcare organizations.



\*\* - Choice of health plan given to eligible by employers or Medicare/Medicaid.

++ - Utilization review/ do DM programs need to be implemented?

## - Outsourcing decision

**Figure 1.1 Disease Management Supply Chain**

## **1.2 Outsourcing in the Health Insurance Industry**

In order to remain financially viable and profitable while adhering and promoting the disease management principles listed above, a medical insurance organization must develop effective strategies for care provision to the affected population [20]. One method for this is the outsourcing of the disease management programs by the Health Management Organization (HMO) to external disease management organizations (DMOs). To date, the decision for outsourcing has been attributed to changes in market costs and not due to internal organization costs. However, internal organization costs have been thought to be just as important to the outsourcing decision as the external market costs, and this was proved empirically in the shipbuilding industry [47]. Our objective is to explain the outsourcing or integration decision of patient care programs based on transaction cost factors, and to determine to what extent that decision is supported, by measuring and comparing costs of the different organizational forms.

We will study the various transaction cost factors as applied to disease management programs, determine the most important ones, that is, the factors which exert the most influence over the outsourcing decision in this industry, and study whether their primary effect is on external market costs or internal organization costs.

## **1.3 Background, Complication and Objectives**

The state of Florida is unique in that it was the first state in the country to develop and implement disease management programs within the state healthcare plans for eligible residents, and encourage the implementation of these plans in the states private HMOs and healthcare providers in the late 1990s. The other states in the country are taking an active interest in the performance of these programs to see if these programs deliver on their promise of reduced healthcare costs, better patient health outcomes and improved efficiency and profitability for the healthcare organizations.

While early research has shown improvements in the health outcomes and costs for patients in the short term, the long-term effects for both the patients and the organizations are not clear and need to be studied further [22, 39].

Employing this econometric analysis to this industry will allow a study of the strategies undertaken by the concerned organizations in order to meet these objectives, and bring out the effect of transaction costs on these organizations, while highlighting the most important transaction cost factors that apply to this particular industry. Thus, it will have an immediate broad impact.

The significance of the proposed research is that it will provide a model that can be widely disseminated and improved upon to assist in the further research and learning in the field of transaction cost economics and factors as applied to disease management programs, medical tasks and the service industry in general. Much will be learned about how internal organization costs influence the outsourcing decision and what transaction cost factors have the greatest influence over the final form of the organization. As various organization costs will also be gathered, this research will also yield valuable information on the costs /savings incurred by the various forms of organization possible in a specific case. Transaction cost analysis applied to the outsourcing of disease management programs will contribute to a deeper understanding of the economics followed by the health management/maintenance organizations (HMOs) and government healthcare entities (Medicare/Medicaid). Objectives of the proposed thesis research are to:

- 1) Determine whether transaction cost analysis can be used to validate the effectiveness of organizational form in disease management programs.
- 2) Identify which of the transaction cost factors (such as asset specificity, uncertainty and complexity) exert greater influence on the outsourcing decisions in disease management programs.
- 3) Determine whether outsourcing leads to fulfillment of disease management objectives.
- 4) To isolate the effects of transactions on the cost of in-house care and outsourcing of care.
- 5) Provide dollar estimates for the costs/savings associated with the sourcing decision.

#### **1.4 Research Approach and Benefits**

In economics and other related disciplines, transaction costs are defined as the costs incurred in addition to the price of the intended economic transaction such as a service, task or product. A number of kinds of transaction cost have come to be known by particular names [38].

Search and information costs are costs such as those incurred in determining that the required good is available on the market, which has the lowest price, etc.

Bargaining costs are the costs required to come to an acceptable agreement with the other party to the transaction, drawing up an appropriate contract and so on. In game theory this is analyzed for instance in the game of chicken.

Policing and enforcement costs are the costs of making sure the other party sticks to the terms of the contract, and taking appropriate action (often through the legal system) if this turns out not to be the case.

The factors that cause transaction costs to be incurred for organizations can be attributed to various factors that can be explained as follows [75, 78, 82, 51, 36, and 61]:

1) Asset Specificity

Williamson [78, 82, and 83] has suggested six main types of asset specificity:

- Site specificity
- Physical asset specificity
- Human asset specificity
- Brand names
- Dedicated assets
- Temporal specificity

2) Uncertainty

3) Similarity/relatedness [47]

4) Frequency

These factors have been explained in the next section. Transaction cost economics can be used as a framework for understanding the healthcare organization's decision to outsource or integrate disease management programs based on these factors. Research in this area has encountered significant difficulty due to the difficulty of obtaining the relevant data, and empirical data have not been applied to the disease management programs, and the evidence of their effectiveness is limited [39, 34]. The application of the above analysis to disease management programs is helpful in explaining the outsourcing protocol followed by many medical organizations. It shows which

factors are the most influential in the decision to outsource patient care, and also helps in providing a dollar estimate of the various organizational forms in this sector, which has not been available before. But most importantly, this research helps in improving profitability for medical organizations, without compromising the aims of the implemented disease management programs, among which are increasing satisfaction and quality of life and reducing costs for the patients.

The benefits of applying the transaction costs analysis to disease management programs are as follows.

- 1) It leads to more effective understanding of the organizational structures of private and government health management/maintenance organizations. In this research, we give explicit attention to the role of internal organization costs in outsourcing decisions. We use transaction cost analysis as a framework to study these costs. Many previous attempts to apply transaction cost economics to various industries have used estimations of reduced form relationship between organizational forms and observed characteristics. Due to this, it was not possible to decipher whether the resulting organizational form was due to changes in market transaction costs or from variations in the costs incurred in organizing the production internally. Using censored regression and the two-stage method outlined below, we can overcome the difficulties generally observed in obtaining direct observations of data, while at the same time giving explicit attention to the role of internal organization costs. Based on this, we can infer whether the effect of a particular variable raises the probability of integration in a particular organization due to increase in the hazards of market exchange or its effects on the internal organization costs.
- 2) It increases understanding of the factors and costs that affect outsourcing. Application of these methods shows which transaction cost factors exert a stronger influence over the outsourcing in the health management organizations, and whether they have a stronger effect on the costs of internal organization or market exchange costs.
- 3) Application of censored regression techniques also leads to the isolation of the effects of attributes of transactions on the cost of organizing within and between firms and provides dollar estimates to these costs. It has been proven that the costs vary systematically with the nature of transaction and that the savings of choosing the right organizational arrangement are substantial [47]. Empirically, it has been shown that in the shipbuilding industry, mistaken

integration of work that is typically outsourced/subcontracted increased internal organization costs by 70%, while outsourcing work normally performed internally within the firm led to organizational costs almost three times those incurred if the jobs were done internally [47]. Transaction cost analysis applied to disease management program outsourcing in the form of censored regression techniques provides a similar estimate of the costs and savings borne by these organizations.

- 4) It also contributes to the research on transaction costs. This work contributes to the research on transaction cost analysis. Various transaction cost factors have been studied in this research. This method of analysis has been applied to both the manufacturing and the construction industry. The factors for scheduling and engineering intensity have been proven to be important in the case of the naval shipbuilding industry [47]. Although the conditions of bounded rationality and opportunism may be universal, the factors that influence them may vary from one industry to another. Hence the effect of the factors considered will be different for different industries, and as a result, it is important for case studies in various other industries be carried out along with more formal empirical analysis. Transaction cost analysis has so far not been applied to the disease management industry and empirical research in this industry using censored regression techniques is yet to be carried out, apart from the analysis and results presented herein. The application of the transaction cost analysis framework to this industry enhances our understanding of health plans decisions regarding outsourcing and their organizational behavior.

## **1.5 Structure of the Report**

This thesis is organized as follows, spanning six chapters.

Chapter 1 introduces the topic area, outlines the reasons for study, and provides details of the research objectives. Chapter 2 focuses on providing an extract of the literature survey prior to forming the hypotheses. Next, chapter 3 summarizes the theoretical concepts and articulates the hypotheses based on the literature review performed on disease management, patient care programs and transaction cost economics. Chapter 4 outlines the research methodology for primary and secondary data collection and analysis that is used for hypothesis testing in our case, and chapter 5 presents the numerical results and inference. Finally, chapter 6 provides the conclusions and the directions for future research in this area.

## **Chapter 2. Literature Review**

### **2.1 Introduction**

Outsourcing is a very well researched topic and considerable research has been done in the field of transaction cost economics to explain the cause and effects of outsourcing in various industries.

This chapter aims to provide a history of transaction cost theory and the previous research on this topic with the help of an extensive literature search involving the study of relevant theoretical concepts and previous related work. The conclusions reached from this exercise have been summarized in chapter 3 to form the background to the work done in this research.

Another objective is to examine the transaction cost theory and disease management literature to find relevant theories and empirical evidence regarding the in-sourcing vs. outsourcing or build vs. buy decision faced by various organizations. The collection of the available results is used in the formulation of the main hypothesis of this research.

### **2.2 Transaction Cost Economics (TCE) Theory**

There is an immense body of literature available in the field of transaction cost economics. A comprehensive review may be found in Shelanski and Klein's [60] 1995 work. The main tenet of transaction cost economics (TCE) suggests that transactions between providers and users of goods or services should be organized in a manner such that transaction costs are minimized.

The theory behind transaction cost analysis was developed by Ronald Coase in his seminal paper, *The Nature of the Firm* (1937) [12], which laid the foundation for all further research done in this area, most notably by Oliver E. Williamson. This theory was used by Coase to develop a theoretical framework for predicting when certain economic tasks would be performed by firms

and when they would be performed on the market, as noted by Robert Kissell and Morton Glantz in *Optimal Trading Strategies*, AMACOM, 2003 [37]. Subsequently, Oliver E. Williamson coined the term transaction cost and has done extensive research in this area, which is elaborated on below.

Organizations and firms usually do not place emphasis on transaction costs. According to Straub and Ang's (1998) [66] research, production cost (which is defined as the amount of money a customer pays the vendor for its services) is given six times more importance than transaction costs. McFetridge and Smith (1989) [49] study outsourcing service contracts in Canada in their research and find that simple production costs are not sufficient to explain the pattern of outsourcing, which validates the theory and effects of transaction costs.

The theory of transaction cost economics focuses on the costs of transactions when a good or service is transferred from a provider to a user. When an organization outsources, the transaction costs will include the costs of searching and selecting the supplier(s), drawing up the contract, performance/results measurement, and dispute resolution (usually involving litigation and/or a third party adjudicator). Conversely, when transactions are internal, the total costs include managing and monitoring costs in addition to the cost of the capital, inputs and raw materials required for the transaction. According to Williamson (1989) [79], the form adopted by the organization, (referred to as governance structure, by Williamson) affects the transaction costs. Transaction costs occur before and after an economic transaction and a central proposition of transaction cost economics is that organizations strive for greater efficiency by implementing governance structures that minimize transaction costs.

Organizations have many options for organizing these transactions via governance structures which vary from spot/open markets for generic goods and services where the buyers and sellers are immaterial to the transaction, to fully vertically integrated organizations, where both buyer and seller can be said to be one and the same and are under joint ownership and control. Between these two extremes of spot markets and vertical integration there are various contracting choices available for the organization to complete its transactions, which include shared ownership of assets or joint ventures.

Williamson (1979, 1981) [76, 77] states that markets are not the best solution for transactions involving asset specificity because buyers and sellers can easily walk, that is, cancel the transaction without any loss to themselves. Markets are also not ideal when one considers the possibility of opportunistic behavior by the parties, which is explained below.

Williamson (1989, 1993) [79, 81] explains opportunistic behavior as follows: the value of the transaction-specific assets in question depends on the continued contract between the buyer and seller, hence, the party that has not invested in these specific assets may be tempted to threaten to walk away from the relationship in order to realize more value from this investment. He also points out that asset specificity plays a major role in the degree of vertical integration and that vertical integration may be the only solution for costly asset specific investments as it is highly difficult for these assets to be transferred or utilized for alternative buyers/sellers and used for other tasks and services.

The types of transaction cost factors have been broadly defined by Leeman (2006) [39] as follows:

- 1) Uncertainty,
- 2) Asset specificity and
- 3) Frequency.

Another factor can be said to be the similarity of the tasks and services in question. Asset specificity is generally regarded as the most crucial transaction cost factor [38]. Others regard Uncertainty to be the most critical factor [76].

### **2.2.1 Uncertainty**

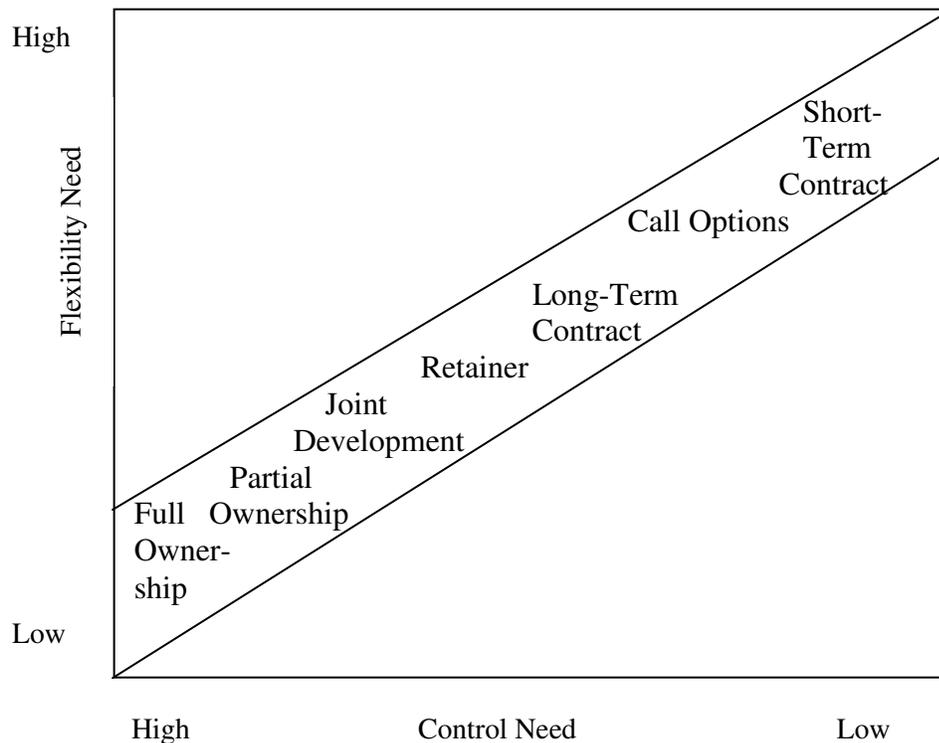
As stated by Leeman [39], “Uncertainty generally refers to how easily performance can be monitored. Monitoring becomes problematic when the task requirements or outcomes are difficult to predict or when the service purchased requires teamwork, making it difficult to connect the product with an individuals input. The greater the uncertainty, the greater the transaction costs incurred in developing and executing a contract in a manner such that all parties are satisfied with the outcome”.

Many researchers have studied the effect of uncertainty on organizational form. Pirrong (1994) [53] found that in ocean shipping the type of contract used (spot markets, medium- and long-term contracts or vertical integration) depends on the uncertainty of providing alternative shipping services for the goods at short notice in the event of a problem or holdup.

Stigler (1951) [65] has theorized that due to uncertainty, industries that are in decline show greater tendency to outsource, whereas the organizations in their growth phase show industries with a greater tendency to integrate. Casson (1986, 1987) [9, 10] studied the shipping industry and has found that that shipping companies running oil tankers and refrigerated cargo ships tend to have ownership of the vessels used for transport early in the company development and are usually leased/contracted in the case of more established companies. This observation supports Stigler's theory given above.

However, a contrasting view to Stigler's theory is available in the literature and can be seen in Harrigan's (1983) [27] research, which has analyzed the vertical integration within 192 firms in 16 different industries in the period between 1960 and 1981. She states that new industries which are inexperienced tend to have less integration and more outsourcing in order to reduce risks. That is, early in an industry's development, when costs and risks are high, firms generally operate with less integration. An example of the computer industry is given, which outsourced microprocessors and memory chips in its infancy, but began to internalize the production of these components as the industry grew and stabilized. The colloquial is stated as the greater prevalence of outsourcing within industries in decline, in order to meet fluctuating demands and market conditions (uncertainty), which can be limited due to high levels of integration. Harrigan also states that certain firms with bargaining power over suppliers, distributors, and customers can reduce prices by reducing supplier profit margins, and can avoid integration, thus the disadvantages associated with it.

These results contrast with Stigler's (1951) [65] hypothesis that firms integrate early during industry development in order to achieve competitive advantages.



**Figure 2.1 Flexibility Need v/s Control Need Option Range**

Source: Quinn, James Brian, and Frederick G. Hilmer, Strategic Outsourcing, Sloan Management Review, Summer 1994, pp. 43–55. [58]

Quinn and Hilmer (1994) [58] have studied uncertainty in terms of flexibility and control. If the firm is subjected to uncertainty in the form of changing demand for its products or services then outsourcing gives the firm the flexibility to meet the changing scenarios but causes it to lose some control over the outsourced activity in terms of execution and performance. They state that when firms outsource they normally transfer certain risks and investments that they would have normally incurred by the contracted party. Figure 2 above shows all the choices of organization form depending on the company's control and flexibility needs.

In order to minimize the effects of uncertainty, Quinn & Hilmer [58] in the same article also suggest that outsourcing be done by carefully taking the firm's skills and resources into account,

and also by comparing the potential of gaining a competitive advantage in the market with all the costs that would be incurred due to contracting.

This compromise has been diagrammatically represented by them as per the matrix in Figure 3. If the activity is such that it allows the organization to gain a competitive advantage while vulnerability/uncertainty is low, then, it can be outsourced, else it should be integrated.

		Degree of strategic vulnerability	
Potential for competitive edge	High	High control needed (integrate)	
		Moderate control needed (special venture or contract arrangement)	
			Low control needed (Outsource)
		High	Low
Low			

**Figure 2.2 Competitive Advantage v/s Strategic Vulnerability Matrix**

Source: Quinn, James Brian, and Frederick G. Hilmer, Strategic Outsourcing, Sloan Management Review, Summer 1994, pp. 43–55. [58]

Both Badaracco (1991) [4] and Harrigan and Newman (1990) [28] state in their research that there is a potential for knowledge leaks when organizations outsource, which, if associated with the source of its competitive advantage can lead the organization to suffer a major setback. These uncertainties or risks, which are the loss of critical skills or loss of control over a supplier, have to be managed by careful monitoring and management of the outsourcing relationship, which leads to an increase in transaction costs for the organization.

The counter argument to this is that transaction costs can still increase due to uncertainty even if the task or service is integrated. Internal employees and departments may fail to perform to their full capacity, and may require policing and monitoring resources to improve performance. In some cases it is more difficult to enforce and measure performance for internal tasks and services than for external suppliers, which increases uncertainty and thus leads to more transaction costs for the organization. Thus, as per Blumberg and Blumberg (1994) [6] an organization may lag behind industry best practices if internal departments are not world-class providers, due to an increase in transaction costs.

Therefore, if the disease management industry were considered to be in a state of growth then previous research evidence would suggest that health plans would tend to vertically integrate to include disease management operations and not outsource it. The disease management sector is indeed seen to be in the growth phase (as per our literature review in the next section), however, it is seen that organizational form for these programs is likely to be outsourced as a result of the specialized nature of this sector.

### **2.2.2 Asset Specificity and its Effect on Organizational Form**

Asset specificity refers to transaction specific investments in human, physical or other forms of capital. Asset specificity also refers to how specifically a particular product or service is designed or produced for a specific customer or if the product or service uses a specific asset. It is broken up into six main types, as explained below:

- 1) Site or location specificity— the location of the buyer and seller in order to economize on inventories or transportation costs, or transportation and inventory costs specific to the transaction;
- 2) Physical asset specificity—investments such as specialized equipment, tools , machines or systems designed for a particular customer or applications;
- 3) Human asset specificity— the skills, experience or knowledge of the people involved in the transaction, or one or both of the parties develop skills or knowledge specific to the buyer-seller relationship;

- 4) Brand specificity—the evaluation and selection of vendors and suppliers based on their reputations, or when the involved parties must maintain the reputation of a shared brand name such as a franchise relationship;
- 5) Dedicated capacity—capacity that is created to serve a particular customer , and this capacity is difficult to adapt to use for alternative customers; and,
- 6) Temporal specificity—the level or importance and specificity of the timing of a particular product or service.

Joskow (1985) [36] has studied asset specificity with respect to mines supplying raw materials to electricity generator plants and his study shows that vertical integration is positively associated with all forms of asset specificity such as site specificity (when transportation costs are high), physical asset specificity, and human capital/know-how specific to the transaction. Stuckey (1983) [67] and Hennart (1988) [32] have researched site specificity specifically and their results support the above results and shows that aluminum refiners generally own their own bauxite mines because of high transportation costs (site specificity) whereas this is not true in the case of tin refiners as the refiners are able to handle different ores.

Masten (1984) [45] studied asset specificity in the aerospace industry and found that integrated components were generally more complex and specialized than ‘buy’ components. The higher transaction costs for these components due to a higher degree of physical and human asset specificity were stated as the causes of integration. Masten, Meehan, and Snyder have extended the above research by studying the organizational form in the U.S. Auto Industry (1989) [46]. They conclude that while physical and site specificity were not the major factors that decide vertical integration, engineering intensity is, and the reason for this is theorized as the greater human asset specificity required for these components and the difficulty of managing this when they are outside the firm make it more suitable for these components and services to be integrated.

Chandler (1961) [11] has also analyzed the maintenance strategies of the North American airline industry after the introduction of jet engines, based on human asset specificity. The airlines had always maintained their own piston engines using their internal maintenance departments. Jet engines were found to require new maintenance skills and facilities but less frequent maintenance. Due to this, the maintenance of these engines continued to be integrated during the

early years of the jet age. However, once maintenance practices and routines became standard across airlines and engine types, internal maintenance departments were outsourced to external independent maintenance specialists

Fuhr and Thorsten [23] have studied the Vertical Governance between Airlines and Airports using transaction cost analysis in 2006. They conclude that temporal specificity and uncertainty play a major role in the contracting between airlines and airports of various sizes.

Anderson and Schmittlein's (1984) [3] work on human asset specificity reinforces the above finding. Their study of the factors that determined the use of company sales staff as opposed to independent distributors led to findings that an internal sales staff is used to reduce transaction costs when the following is required: 1) specialized training, 2) detailed or proprietary knowledge of the selling company 3) continuing relationship between salespersons and clients 4) detailed knowledge of product or customer, and 5) when output measures of sales staff are unreliable.

Masten, Meehan, and Snyder [47] have also studied the organizational form and associated costs in their 1991 study of naval shipbuilding industry and found that higher the importance of timely completion/scheduling of the component in construction the higher is the likelihood of integration. This is because an interruption in any stage of construction disrupts all subsequent operations by having a cascading effect which causes delays to the whole project. This also gives subcontractors incentives to delay in order to gain price concessions, which is a type of opportunistic behavior.

Monteverde and Teece (1982a) [50] found that in General Motors and Ford the probability that a component is produced in-house increased with the engineering effort required to design it. This has been attributed to human capital specificity due to the engineering knowledge required in these applications. Monteverde and Teece's work (1982b) [51] on physical specificity found that automobile manufacturers in general were more likely to retain title to the more specialized and expensive tooling used by suppliers. This again supports the theory that greater the asset specificity, the higher the incentive is for organizations to integrate those tasks/applications.

Lehmann and O'Shaughnessy (1974) [40] have studied Reputation (Brand Specificity) and they state that it is very important in the selection and evaluation of vendors and suppliers.

They say that this is due to management's desire to reduce risks to their companies and for themselves. This is achieved by selecting suppliers with a good reputation and high credibility, which can also improve the image of the contracting firm itself in some cases.

Panayides and Cullinane (2002) [52] have studied the importance of reputation in ship manager selection. Their research was aimed at finding the most important criteria for ship manager evaluation and selection. Their sample size consisted of 48 ship management companies and 36 ship owners. They state that the inspection for selection is done mainly on two levels; the first level is financial variables, profitability, location and managerial ability. The second and more important level is a measure of the manager's reputation, image and reliability, integrity, trustworthiness, and commitment. Thus, they state that brand specificity in the form of the ship manager's experience, establishment and status is a significant factor for the organizational form chosen by shipping companies for ship management. Reputation of the contracting parties can be said to reduce the risk of opportunism, which would reduce the monitoring costs and increase the efficiency and profits to both parties involved.

The above literature shows that in general, an increase in asset specificity in any form is positively related to integration.

### **2.2.3 Similarity and Frequency**

Similarity can be said to refer to the nature of the tasks or processes and how closely they resemble the ones done on a regular basis by the firm or organization. Leeman (2006) [39] further states that transaction cost analysis studies the relationship between characteristics of transactions and the forms of governance organizations implement to negotiate and execute those transactions. Some examples of organizational structure include long term contracting, short-term contracting and internal production. The view shared by most economists is that organizations choose specific arrangements by comparing the costs of transacting under each. This insight needed empirical support, which was provided by noting the observable attributes of transactions by Williamson (1975, 1979) and Klein et al. [75, 76, 60].

However, these efforts have generally concentrated on factors aggravating the hazards of market exchange, and the costs of internal organization have been treated only as a barrier to be overcome before integration (Masten et al., 1991) [47].

Frequency refers to how often the purchaser transacts in the market. Due to economies of scale, frequency decreases the per transaction cost of asset-specific investments. Therefore, greater transaction frequency will enhance the value of asset-specific investments, for example, the costs of implementing new care management processes (Leeman, 2006) [39].

Masten et al. [47] have given specific attention to the role of transaction cost factors on internal organization costs in organization integration decisions, and have provided the empirical study of a naval project, which has provided dollar estimates of the costs of various organizational arrangements. The most important result of this research is regarding the contribution of changes in market and internal organization costs to the final arrangement adopted by the firm. It is well known that internal organization sacrifices the advantages of market exchange, while preventing problems such as opportunism, scheduling and uncertainty. However, this demands greater investments in administration and monitoring (Williamson, 1985; 1990) [78, 80].

Economists and theorists have paid little attention to the influence that these factors make on the costs of managing and monitoring tasks and services internally, and to what extent they weigh on the form finally adopted by the organization. They have concentrated on how these factors affect the market prices, while neglecting the former effect. Ronald Coase has been one of the exceptions to this view and states “the effect of activities in which a firm is already engaged on the cost of undertaking additional activities” is essential to explaining why particular operations are chosen within specific firms (1988:40) [14]. He goes on to say, “The way in which industry is organized is...dependent on the relation between the costs of carrying out transactions on the market and the costs of organizing an activity within that firm which can perform this task at lowest costs. Furthermore, the costs of organizing an activity within any given firm depends on what other activities it is engaged in. A given set of activities will facilitate the carrying out of some activities, but hinder the performance of others. It is these relationships which determine the actual organization of industry.” (1972:64) [13]. He also states that internal organization costs are likely to be higher for transactions other than those in which the firm is already engaged in, for which there is a higher degree of uncertainty. Asset specificities of the various types explained above tend to raise organization costs, if integration is carried out, and also raises market exchange costs, if outsourcing is favored, however, in the case of this factor, integration is usually preferred as it allows greater flexibility for change and modifications. Similarly, uncertainty and complexity, while producing a net increase in market as well as internal organization costs, favors

integration to subcontracting as integration gives the organization allows the organization to adapt to changing situations and circumstances, where outsourcing does not. The similarity of transactions, on the other hand, is unlikely to drive down market costs as the parties engaged in the bargaining are most concerned about the final outcomes and not the manner in which the goods or services are provided [47]. In order to verify the above statements, empirical data needs to be collected in order to support or refute them. This has only been done in the shipbuilding industry and needs to be applied to the disease management field in order to study the effects of these factors on disease management program sourcing decision.

#### **2.2.4 Bounded Rationality**

Managers and organizations have limited managerial time and control, and hence they cannot manage all tasks internally or plan and contract for all possibilities in the future in the case of outsourced tasks or services. This is due to bounded rationality, and thus, bounded rationality influences organizations in their attempts to reduce transaction costs.

The theory of bounded rationality was proposed in 1957 by Herbert Simon [62, 63], and it can be explained as the limitations on decision-making due by time, costs, human abilities, availability of information, and technology. He states, “Bounded rationality is a central theme in behavioral economics. It is concerned with the ways in which the actual decision-making process influences decisions. Theories of bounded rationality relax one or more assumptions of standard expected utility theory”.

For most transactions, markets are the preferred governance structure as markets provide the incentives to cut costs and maximize value net of production costs, while at the same time they allow the parties involved to respond quickly to changes in the market. As stated before by Williamson (1981) [77], markets are not the ideal solution for transactions involving asset specificity because buyers and sellers can cancel the transaction entirely. Contracts of differing lengths can offer some protection against the drawbacks of market transactions; however bounded rationality makes it impossible to draw up contracts that cover all possible circumstances, due to which the involved parties may indulge in opportunistic behavior to make profits.

As a result, complex internal control and monitoring systems may be needed to police the contract, make changes, and settle disputes if needed. Thus, bounded rationality brings out the negative aspects of market transactions due to the possibility of opportunistic behavior by the parties and increases transaction costs. For other types of transactions a more integrated governance structure may be desired. If outsourcing is not possible due to asset specificity, and bounded rationality, vertical integration can be used in order to maximize profits and reduce transaction costs. Bounded rationality also plays a role in internal organization as control within the organization may be lacking in certain aspects, due to which opportunism by employees and an increase in transaction costs within the firm may be seen. Therefore bounded rationality affects transaction costs in both governance structures, and organizational form should be chosen in order to minimize it.

### **2.2.5 Core Competence and Transaction Costs**

Prahalad and Hamel [54] introduced the concept of core competence in their 1990 study which they define as “the collective learning in the organization, especially in coordinating diverse production skills and integrating multiple streams of technologies.” Excellence in a few core competencies is what gives the organization a competitive edge in the market.

Quinn and Hilmer (1994) [58] in their article Strategic Outsourcing recommend outsourcing only non-core activities to minimize transaction costs. This suggestion has been made so that firms can concentrate their limited internal resources on a set of core competencies and tasks where they can achieve pre-eminence and provide unique value for their customers. In order to differentiate and identify these core functions in an organization, they have put forth the guidelines given below:

- 1) Core competencies are limited in number.
- 2) Core competencies are flexible and long-term platforms capable of change.
- 3) Core competencies are skills or knowledge sets not products or functions. They also cut across traditional functions. Hence, they are activities that are based on knowledge rather than on ownership of assets.
- 4) Core competencies should be embedded in the organizations systems and not dependent on a few people.

- 5) Core competencies can be used as sources of leverage in the value chain.
- 6) Core competencies are Core functions/services that are important to the customer in the long term such as understanding and serving the customer.

We can infer from the above that companies must retain only activities that give them the competitive advantage and other tasks and services may be outsourced. However, Quinn & Hilmer [58] point out that this is in fact not possible as the “supplier markets are not totally reliable and efficient”. According to them most outsourcing will entail some risks, which have been elaborated above.

Harrigan [27] supports the above recommendation with her 1983 work in which she analyzes the vertical integration strategies of 192 firms in 16 different industries from 1960 to 1981. She finds that generally firms internalize the tasks and services that they consider to be their core competencies or those that contribute to their competitive advantage in order to minimize transaction costs. One example cited in this work is how computer firms manufactured the logic chips and processors for their product internally but purchased the other components. Another example of this is that pharmaceutical firms used their own trained sales agents for marketing their medical products in order to protect their patents and increase sales and also integrated production of certain chemicals and pharmaceuticals during high demand.

Thus we see that evidence from the business management literature shows that integration reduces transaction costs associated with market transactions and common administrative functions.

However, integration can also increase transaction costs in the form of internal coordination, policing and enforcement costs and reduce incentives to maximize performance and efficiency within the organization. Hence it is clear from the above that firms will benefit more by firstly outsourcing activities or parts of activities that are less critical to its survival.

The main benefits of outsourcing can be summarized as stated by Corbett (1995) [15], and they are: improved business focus, access to world-class capabilities, reduced cycle times and improved quality, sharing risks and costs in new technology. Other main benefits are reducing

operating costs, converting capital investment in non-core functions into operating expense and gaining better control integrated tasks.

Buzzel (1983) [8] has studied 1649 manufacturing units from the Profit Impact of Market Strategy (PIMS) database. His research shows that either a very high or a very low level of vertical integration yields an above average rate of return while earnings are lowest in the middle, and he recommends vertical integration only when a company needs savings as well as high control over its tasks and services. A measure of integration is given by the value added to sales ratio. According to him, the advantages of in-sourcing are lower transaction costs, supply assurances, improved coordination, and lower uncertainty. The disadvantages are capital investments, unbalanced throughput, reduced flexibility and a loss of specialization.

D'Aveni and Ravenscraft's (1994) [16] work on the benefits of vertical integration support Buzzel's findings by showing that vertical integration can reduce total costs by avoiding the transaction costs associated with market transactions, combining administrative functions previously performed separately, and providing better information about costs. Like Buzzel, they also point out that integration can increase transaction costs in the form of costs for coordination and production. The additional coordination of activities required in integrated organizations may increase overhead. They state that production costs may increase because of the lack of market pressure to improve the efficiency of internal processes and employees, lower economies of scale, or failure to innovate. Other costs that increase are the costs needed to monitor market information, manage inventories, and plan and schedule activities.

In closing, we see that higher asset specificity in all its forms is generally associated with greater vertical integration due to higher transaction costs of outsourcing these specific tasks. Other factors that influence the transaction costs and organizational form are bounded rationality, frequency and uncertainty. Thus, in making decisions regarding organizational form, it is important to consider not only the actual cost of the good or service in each case, but also the transaction cost factors that will be most prominent in that scenario, and the level of these costs when managing the transaction both internally and externally.

## 2.2.6 Empirical Measurement of Transaction Costs

To the authors knowledge based on the literature review, the empirical estimation of costs incurred due to transaction cost factors has been done only twice before, first by Wallis and North [73] in 1986, who attempted to measure transaction costs of the economy over a 100 years. However, they faced severe problems in defining and more so in measuring transaction costs that are detailed in the methodology section. Their analysis concludes that the transaction sector is a significant part of the economy and grew from 25% to 40 % between 1870 and 1970.

The second is by Masten , Meehan and Snyder in 1991 [47], who provided empirical evidence of both the influence of each transaction cost factor but also provided cost estimates of each organizational form applicable to a shipyard involved in Naval construction projects. The problems faced by Wallis and North were mitigated by using switching regression techniques. This was done using probit regression models to compute the effects of each factor on the form actually adopted. A problem of selection bias was encountered for the second stage cost calculation since the efficient organization structure is chosen, the other forms are not observed, for which the Heckman two-step procedure and correction factor was used to eliminate the selection bias as outlined below. The structural equations were estimated as censored regression models analogous to the way actual and reservation wages are estimated in labor supply applications. From this technique they obtain actual dollar estimates of transaction costs and can therefore estimate the magnitude of individual coefficients and not just their relative impact. They found that transaction costs account for 14% of the total value of all components analyzed and that costs of the components made internally would rise to three times the actual were they to be outsourced. Integration of the contracted components would lead to a 70% increase in transaction costs. Masten [44] has applied the above methodology to assess the performance implications of governance choices and its effect on business performance.

Also reported is the fact that the transaction cost factors mainly affect the costs if internal organization, rather than market costs as is normally assumed. Thus, the importance of organizational form is substantial.

### **2.3 Disease Management Literature Review**

Disease management is quickly emerging as one of the most important new areas of medical management as noted by Quilty and Lewis in the article Case Studies in Disease Management in *Medical Interface Magazine* [56]. The World Health Organization (WHO) estimates that chronic diseases make up 60% of the global disease burden, which is expected to rise to 80% by the year 2020 for developing countries [55]. For the United States, the Centers for Disease Control and Prevention (CDC) estimate the total cost for diabetes as \$137.7 billion in 1995. According to Thompson, Edelsberg, Kinsey, and Oster [70], nearly half of the American workforce is either overweight or obese. Americans with chronic conditions account for 75% of total healthcare costs [33]. Chronic illnesses are the major cause of morbidity in the United States, and due to the increase of senior population in the country, the prevalence of these conditions is bound to increase. At the same time the effectiveness of the U.S. healthcare system in providing care for this segment of the population has been wanting. The medical system has also come under harsh criticism for rising costs and large deviation from best care practices, which is to say the treatment and care which is most suitable for the affected person at that particular time. As noted by Wheatley (2002) [74],” amid rising healthcare expenditures and declining tax revenues state efforts to expand access to health insurance coverage have been put on hold in many parts of the country. Recently, states have had to take a number of difficult steps to reduce program expenditures, including restricting eligibility, reducing benefits, and cutting provider payments. These measures generate cost savings but also restrict access to care. Another option, which is now being more widely adopted by states, is to develop disease management (DM) programs that are designed to contain costs by improving health among the chronically ill. Disease management programs are meant to benefit both the medical insurance organization and the consumer/patient by containing costs by improving health among the chronically ill. More than 20 states are now engaged in developing and implementing disease management programs for their primary care case management and fee for service populations [75]. The popularity of disease management springs from the fact that the proactive management or prevention of chronic conditions presents the single largest opportunity to improve health and reduce healthcare costs.

Disease Management programs focus on patient identification, monitoring and early intervention. This shifts healthcare expenses to less invasive and expensive care, thus, disease management programs are meant to strive to achieve two seemingly conflicting goals: improving health care

while achieving cost savings at the same time. These programs work by drawing on the commitment and self-interest of patients, expert coaching, monitoring and treatment by experienced nurses. The treatment guidelines are grounded in evidence-based medicine. These resources are deployed to monitor patients' conditions and coordinate treatments with the physicians in various settings and diseases.

According to Lewis [42], in disease management, the word intervention can loosely be defined as “that set of products, services, education, expert resources and data offered to the patient, patient’s family/caregiver, and/or provider in order to reduce the likelihood of acute exacerbations and complications and/or to improve the baseline health status of the member overall.” The term intervention is also used in medicine generally, but is done so to define the medical treatments given to a patient. The main difference between a medical intervention and one through a disease management program is one of duration. A medical intervention is usually a treatment, procedure, medical test or therapy, a disease management intervention consists of patient monitoring, follow-through, support and assistance and outcome reporting. Interventions done through these programs can have many points of contact and changes as per the condition of the individual, and may last a lifetime [42].

These programs tackle critical factors that have the greatest influence on quality of life, health and associated costs for most of the populations, especially the chronically ill segment. Currently many DMOs and health plans have overhauled their programs to manage co-morbid patients, i.e. patients with two or more chronic conditions.

### **2.3.1 A Brief History of Disease Management**

According to Boston Consulting Group (BCG) [48]; the earliest known implementation of disease management was the launch of blood glucose monitoring (BGM) units to diabetes patients in the 1980s as this required significant education and monitoring of patients along with the setup of the required infrastructure, and the mindset of both patients and doctors needed to be modified.

This was followed by the first wave of DM programs in the early 1990s, mainly supported by pharmaceutical companies. Pharmaceutical companies supported these programs as they knew that prescription drugs help keep diseases in check and would reduce or minimize hospitalization, particularly in chronic conditions. Many health plans were skeptical, as they view it as a ploy to

sell more drugs. Another limitation they had was that these first-generation pharmaceutical company-sponsored programs came with too many formulary constraints [57]. Thus, most of these programs had closed by the end of the decade.

A second wave started in the middle of the decade when entrepreneurs began to work to serve the large demand for disease management services, which required specialized technology, data mining and management. These early DMOs usually focused on a single disease at a time, and recently there has been a change towards managing co-morbidities, especially in the case of Medicare and Medicaid.

The latest and current wave of disease management has been fuelled by the health plans as they have widely embraced these programs, and support and provide disease management programs either internally or through contracting with external vendors. Today many health plans are working to integrate these programs into other aspects of medical management such as wellness programs. The accreditation by the National Committee for Quality Assurance (NCQA) has helped in wider acceptance of these programs as well.

### **2.3.2 Current State of the Disease Management Industry**

According to the BCG report *Realizing the Promise of Disease Management* [48], published in Feb 2006, today DM enjoys widespread use amongst the majority of U.S. health plans.

According to the above report, out of the 120 health plans assessed from the 150 total in the U.S, all but 4 offered DM programs, meaning that 96% of the American health plans survey offered disease management programs.

DM is now viewed as a competitive necessity according to more than 80% of the decision makers in the health plans studied by BCG in the study noted above. 72 of the 120 health plans surveyed stated cost savings as their reason for disease management program implementation. DM programs are widespread today even though there is marked uncertainty about results, savings and outcomes measurement methodology, and DM vendors or disease management organizations (DMOs) have enjoyed rapid growth over the last decade.

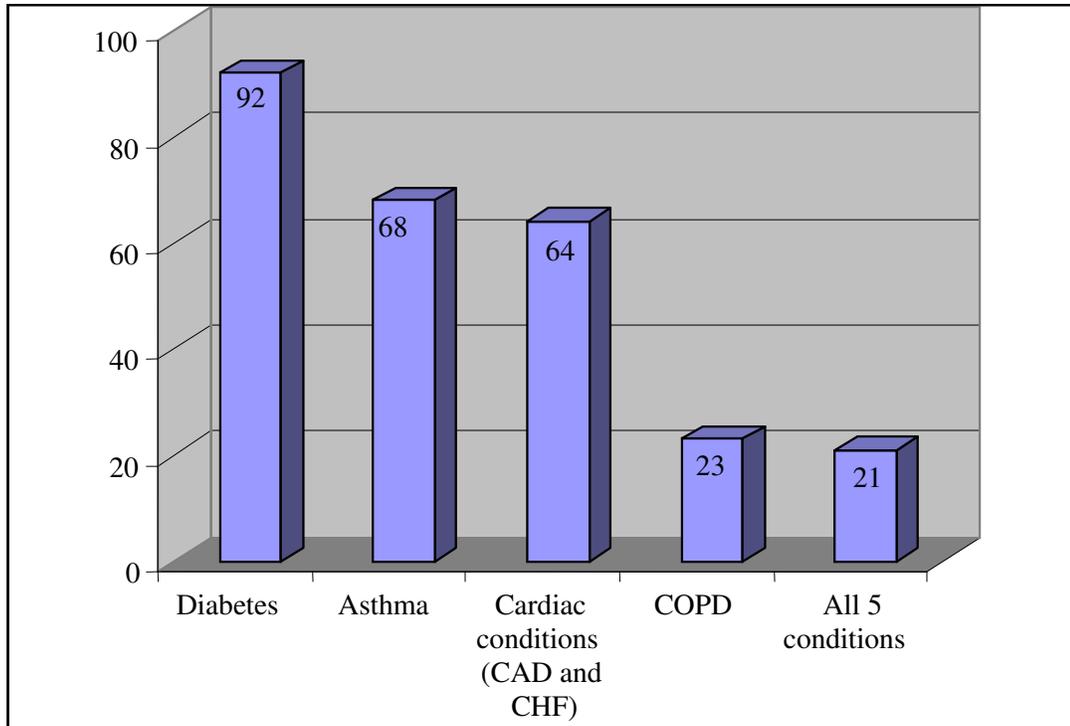
The Disease Management Purchasing Consortium [19] estimates that DMO revenues have increased from \$78 million in 1997 to almost \$1.2 billion in 2005, which gives a compound annual growth rate (CAGR) of 40%. The revenue is expected to grow to \$1.8 billion in 2008, with the growth coming from Medicare and Medicaid. Today many businesses offer DM services, and most DMOs have expanded beyond their focus on a single disease. Humana is generally acknowledged as the leader in disease management [71]. Amongst DMOs, the top five based on market revenues are: Healthways, Health Dialog, CorSolutions, LifeMasters Supported Healthcare, and Matria Healthcare. DMOs have also diversified into informatics, where they sell data and analysis tools for employers to allow them to assess their employee health and health plan performance (Source- DMPC) [71].

The BCG report Realizing the Promise of Disease Management [48] finds that health plans are almost equally as likely to develop and run these DM programs internally as they are to contract with external DMOs to purchase DM services, given the situation within the organization and associated transaction costs. Another option available to them is the combination or hybrid approach, where some health plans combine internal and external resources- such as in-house nurses and purchased software in order to execute DM programs.

Although private U.S health plans are the largest implementers of DM programs today and majority of employers access DM through these health plans, several other sectors such as the direct-to-employer segment is a rising trend and this segment is growing rapidly. Employers are also taking an active interest in managing and coordinating employee health plans and disease management programs and frequently request it when contracting with a health plan. Large employers are also likely to contract separately with a DMO for disease management programs separately from their health plan. These employers usually have multiple health plans and seek a single DM benefit that they can apply across the organization for their employees. For large employers, disease management is growing in importance because they increasingly see the value of such programs in reducing absenteeism and short-term disability expense, not to mention employee morale and retention [2].

Federal and state governments are also getting heavily involved in DM using the Centers for Medicare and Medicaid (CMS) pilots to implement these programs. In addition, Governments abroad are showing an increased interest in this sector. Given the above, we see that disease

management programs are usually implemented for diabetes, asthma, coronary artery disease (CAD), congestive heart failure (CHF) and chronic obstructive pulmonary disease (COPD), which are known as the five core chronic diseases. The number of health plans that offer all five programs represent only 21% of the total, as can be seen from figure 2.3. Also, the number of health plans offering these programs for other chronic diseases such as end-stage renal disease, lower-back pain and cancer are low.



**Figure 2.3 Percentage of DM Programs Offered by Health Plans in the U.S.**

Source: BCG Landscape database, Feb 2006. [48]

The highest governing body overseeing all the organizations in the U.S is the Disease Management Association of America (DMAA) [19]. DMAA is a non-profit association that represents all stakeholders in the DM community. The association does this through public and private advocacy by targeting the healthcare industry, government agencies, employers, and the general public to educate them on the important role DM programs play in improving healthcare quality and outcomes for chronically ill patients [21]. The components of disease management as defined by the DMAA [22] are:

- 1) Population identification processes;
- 2) Evidence-based practice guidelines;
- 3) Collaborative practice models to include physician and support-service providers;
- 4) Patient self-management education (may include primary prevention, behavior modification programs, and compliance/surveillance);
- 5) Process and outcomes measurement, evaluation, and management;
- 6) Routine reporting/feedback loop (may include communication with patient, physician, health plan and ancillary providers, and practice profiling).

Full-service disease management programs are those that include all six components. Programs consisting of fewer components are known as disease management support services.

Traditionally, disease management has focused on the big five chronic diseases: ischemic heart disease, diabetes, COPD, asthma and heart failure. Disease management programs generally are offered telephonically, involving interaction with a trained nursing professional, and require an extended series of interactions, including a strong educational element. Patients are expected to play an active role in managing their diseases. Because of the presence of co-morbidities or multiple conditions in most high-risk patients, this approach may become operationally difficult to execute, with patients being cared for by more than one program. Over time, the industry has moved more toward a whole person model in which all the diseases a patient has are managed by a single disease management program (Source-DMAA) [18].

As stated by the Disease Management Purchasing Consortium (DMPC) [19], disease management requires a comprehensive clinical and economic understanding of a disease state that can only be developed through a team approach. Clinical input is required to design the interventions, identify patients, and understand the impact of co-morbidities. Information systems input is required to integrate the disparate data bases of medical information for a particular disease. Legal and network development assistance is required for contracting and to understand how disease costs are impacted by capitation arrangements. And once a program is developed marketing support will be required to develop physician communication materials. According to Managed Care magazine [43], a typical disease management program consists of the following teams:

Program administrators: These are the individuals who run a health plan and are the best source of information on organizational structure, goals and expectations, pay and incentive programs, and fiscal commitments to the disease management program. The success of such programs is dependent on the support it receives from the administration during its development and implementation.

Pharmacists: Academic and professional training in pharmacotherapeutics and pharmaceutical care empowers pharmacists to play a critical role in disease management. Pharmacists in highly integrated managed care settings participate in formulary decisions, drug treatment protocols and critical pathway design. Pharmacists in disease management programs also perform the following activities –

- 1) Patient education concerning drug use, especially in high-risk/high-use cases.
- 2) Compliance education and monitoring for selected populations.
- 3) Disease state monitoring (blood glucose, blood pressure, serum cholesterol, etc.).
- 4) General wellness education.
- 5) Intervention with physicians to encourage drug protocol adherence.

Information managers: Data analysis plays a critical role in designing and operating a DM program. For the implementation of these programs, algorithms, based on specific correlates of drug, diagnosis, procedure and specialist codes are needed, to query claims data in order to identify the health plan's members who have the diseases in question. As a result of this level of specificity, the entire population with these diseases can be identified. Baseline measurements are necessary for later comparisons to ascertain whether care has been improved and costs have been controlled. Information managers help the planning team decide on data formats and definitions. They determine the usefulness of current information systems and also promote exchange of appropriate data elements among the partners. Continual improvement of the information systems used in the programs is necessary in order to capture and track data used in outcomes research and the information required for future improvement.

Finance managers: The programs finance team is needed to analyze current costs of care, including the costs of failing to achieve intended outcomes and the predicted financial consequences of the disease management program. In addition, this team is responsible for

negotiating contracts among the disease state management partners, and for clarifying arrangements among them with regard to risk sharing and capitation.

Florida operates the largest (and one of the oldest) Medicaid disease management programs in the country, which was initiated in 1998. Florida has the fourth largest Medicaid population in the nation, with 2.1 million eligibles and \$8.8 billion spending in FY 00-01; \$9.9 billion appropriations for FY 01-02; \$ 11 billion FY 02-03, and \$13 billion in FY 04-05 [17]. The Florida disease management program is the most comprehensive disease management program in the nation for Medicaid recipients [68, 1]. The diseases covered by Medicaid DM programs are asthma, HIV/AIDS, CHF, hemophilia, ESRD, diabetes, hypertension, pre-diabetes and depression. In May 2001, a Florida legislative audit was released which criticized the DM program for not being close to producing the projected savings of \$113 million over the period of 1998 to 2001 as was initially expected. It has also been found that while the DM programs generally reduced inpatient hospital costs, produced improvements in patient care quality and led to a reduction in spending, these reductions were generally offset by DM program costs [75]. Table 2.1 shows the most popular disease management programs in the country, while table 2.2 reports the tools used for their implementation.

**Table 2.1 Disease Management Program Statistics Across the U.S.**

Source: Managed Healthcare Executive; Apr 2006. [72]

Disease management programs offered by HMOs:

Disease state	Percentage of HMOs offering programs
Diabetes	81.5%
Asthma	79.6%
Cardiovascular disorders	64.7%
High – risk pregnancy	31.4%
Hypertension	20.0%
COPD(chronic obstructive pulmonary disorder)	16.6%
Multiple sclerosis	8.4%
HIV/AIDS	7.7%
Gastrointestinal disorders	4.3%
Hormonal therapy	1.9%

Other disease management programs offered	45.6%
Top three other programs offered:	
Low – back pain	32.1%
Smoking cessation	22.6%
ESRD (end stage renal disease)	22.2%

**Table 2.2 Method of Disease Management Program Implementation**

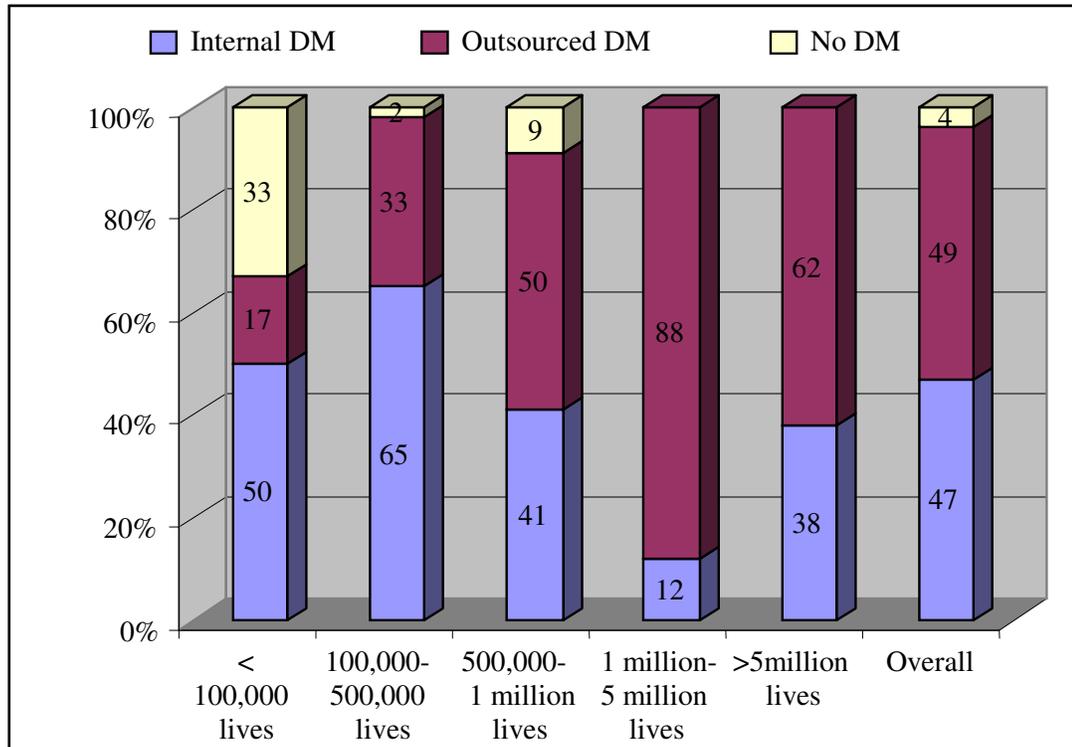
Source: Tracy Walker, Managed Healthcare Executive; Apr 2006. [72]

Implementation tool	Percentage of HMOS offering service
Patient education newsletter	71.4%
Physician education newsletters	59.6%
Information on web site	54.0%
Patient education courses	48.4%
Reminders at physician visits	38.5%
Physician education courses	34.8%

The investment required in terms of capital and human resources is thus significant when it comes to the implementation of disease management programs. A healthcare organization has the choice of implementing such programs itself, or contracting them to outside disease management vendors. In order to remain profitable and financially viable while upholding the principles of disease management and reducing healthcare costs, a medical care provider must develop effective strategies, as noted by Einstein [20]. Outsourcing of these programs to disease management organizations (DMOs) is one strategy that is widely practiced.

### **2.3.3 Effect of Transaction Cost Factors on DM Organizational Form**

We can see the effect of TCE factors on health plans in the survey conducted by BCG in February 2006 [48]. As seen in figure 2.4; health plans are as likely to integrate DM programs as they are to outsource them to a DMO. One way larger health plans have integrated their disease management programs is by purchasing the DMO outright. For example, Wellpoint has purchased Health Management Corporation and UnitedHealth Group has purchased the DMO Optum. Health plans such as Cigna have contracted with DMOs, while others such as Kaiser Permanente have a completely integrated approach. The decision on organizational form, according to the February 2006 BCG report, is made at an individual level by each health plan. They state “it’s not the payer’s size but the perspective of senior management that largely determines whether the payer develops its own DM programs or turns to the market for external options.”



**Figure 2.4 Organizational Form of DM Programs in Health Plans**

Source: Realizing the promise of disease management, BCG, 2006. [48]

The BCG report continues, “they (health plans) recognize the capabilities required to implement the approach and the difficulties involved”, which can be interpreted as asset specificity in the form of technology and software. Also, they go on to say, “[Health plans] view disease management as a highly specialized set of skills that are difficult to master or replicate at low cost. Some payers may view disease management as so central to their business that they will make every effort to make or bring the approach in-house. Others may feel that they cannot afford the fees associated with outsourcing or they can best limit their expenditures by relying on an internal or assembled program” [48]. This shows that human and physical asset specificity plays a large part in determining the organizational form for these programs.

There is uncertainty regarding the savings for the health plan with these programs, and the savings for different programs can be realized at different times.

For example, end-stage renal disease savings can begin in as few as 45 days [59], whereas a Healthcare Business roundtable consensus showed an average interval before savings are realized to be approximately 18 months for other programs [29].

In addition, the BCG report shows that due to lack of a standard methodology in order to measure effects and outcomes, health plans face uncertainty in terms of measuring results, and by choosing an appropriate organizational form, they try to minimize the transaction costs associated with this uncertainty. The biggest obstacle in the path of disease management is that no standard methodology exists for measuring savings and outcomes. The results reporting done for disease management programs consists of usually three outcome measures—process outcomes, i.e. (Did the compliance rate go up?), health status outcomes, i.e. (Did ER visits decline? Did self-reported health assessment scores improve?), and member satisfaction. However, due to no set standards across the industry, the methods used to measure these usually vary from organization to organization. A common mistake is the first is regression to the mean. Any disease management program which starts with last year's high users—a common starting point in asthma and CHF disease management—will automatically show improvement simply because few diseases progress linearly.

Although various industry groups such as the Disease Management Association of America (DMAA) and the DMPC [24] have issued guidelines, there has not been an agreement in terms of adopting a particular methodology, which introduces uncertainty and increases the transaction costs of implementing the program in each organizational form as “each payer will need to examine a variety of issues, such as the magnitude and reliability of its savings measurements”, and “we expect other payers to find disease management so resource intensive and difficult to manage effectively that they will turn to DMOs when their serviced-delivery or internal outcomes prove unsatisfactory” [48]. Uncertainty is also stated as the risk of failure for a disease management program implemented by a health plan, which would cause a setback to the company. Additionally, health plans are seen to look for “common vision and committed leadership” while searching for an appropriate DMO, which is an example of brand specificity. Another view of brand specificity is given in the DMPC report *Outsourcing: Lessons Learned* as “examples of favorable first contracts would be NYLCare-AirLogix, Foundation-Vivra Specialty Partners, Humana-Ralin, Humana-Paidos, Humana-Baxter, Principal-Accordant, and a large

number of health plans fortunate enough to receive programs that were literally given away, no strings attached, in order for a vendor to start generating experience and outcomes.”

On the surface, it might seem risky and problematic to contract with a new or inexperienced vendor for disease management programs. However, due to the relatively new nature of the industry and the unique requirements it entails, new/inexperienced vendors have actually shown better performance as compared to established vendors/DMOs, as noted by the DMPC. As an example, Apria was an established DMO with a vast experience in asthma, and Stuart Disease Management Services (financed by Zeneca), were handling programs for various national health plans, but both pulled out of disease management and left their customers (the health plans with which they were contracted) in the lurch with what are now essentially orphan disease management programs [41]. The above shows the pitfalls of stressing on brand name and reputation and its effect on transaction costs to the level that the programs failed.

Frequency is reported as the number of interventions as well as the retention and penetration among its customers by the DM program. Most health plans screen all policy holders for program eligibility using their preferred algorithms, which take into account the medical history and risk of the individual. If eligible, the individual is enrolled into the program at no expense or for a small monthly fee. The person is free to opt out of the program at any time. Due to this, the number of people enrolled in a particular program is always in flux. The adherence of the patients to the program protocol is also something that needs to be constantly monitored and hence the frequency of contact within a program can vary significantly based on the characteristics of the people enrolled. Hence, this factor also plays an important role in the final form adopted by the health plan for these programs.

According to the DMPC report *Outsourcing: Disease Management's Magic Bullet* (1999) [42], Outsourcing is not always the answer for health plans any more than building programs internally is always the answer. Many health plan medical directors are given directives along the lines of: “You have to institute a disease management program, and you have to do it within your existing budget.” [57]. Thus, there are many variables that influence internalizing or contracting a particular program in order to maximize benefits and profits.

From the above, three main factors can be used to distinguish between those disease categories and health plan circumstances which should lead to a buy decision and those which should lead to a build decision [42]:

- 1) Health plan organization, culture, and budget
- 2) Severity of disease - disease management programs which look like a typical health plan's day-to-day operations can be successfully built by most health plans, but those which require a set of skills not normally found within a health plan are better served through outsourcing to an expert vendor. This shows how similarity may affect the organization form of these programs in health plans.
- 3) Availability of tools and expertise - The more widespread the expertise and tools available for patient management in a particular category, the easier it is to build a program. For instance, health plans often build their own prenatal care programs, using readily available scripts to help their call center nurses triage pregnant members to identify those needing the most attention. The experience base in pregnancy management is built on close to 4,000,000 US pregnancies every year. Rare diseases by definition lack that experience base, and hence expertise and tools are much harder to find. For instance, the nationwide experience base for hemophilia is built on only 20,000 patients. In the case of rare diseases, a health plan can spend more time just trying to assemble the requisite tools itself (assembling the tools being a small piece of the overall disease management program) than it would spend creating an entire program through an outsource. However, Evaluating and selecting vendors, contracting, and claims analysis require some effort and expertise. If integrated, a health plan would need to purchase its own retrospective claims analysis/predictive utilization software. Such a tool can help identify tomorrow's high users (the people one wants in a disease management program) as well as ones from previous periods. Such software, such as CodeReview, is helpful but not exhaustive. Several vendors have very sophisticated algorithms, supervised by medical directors, to find opportunities which software alone can overlook, and they guarantee significant amounts of savings.

The above statements again show how asset specificity in the form of physical and human asset specificity affect the form chosen by health plans for these programs.

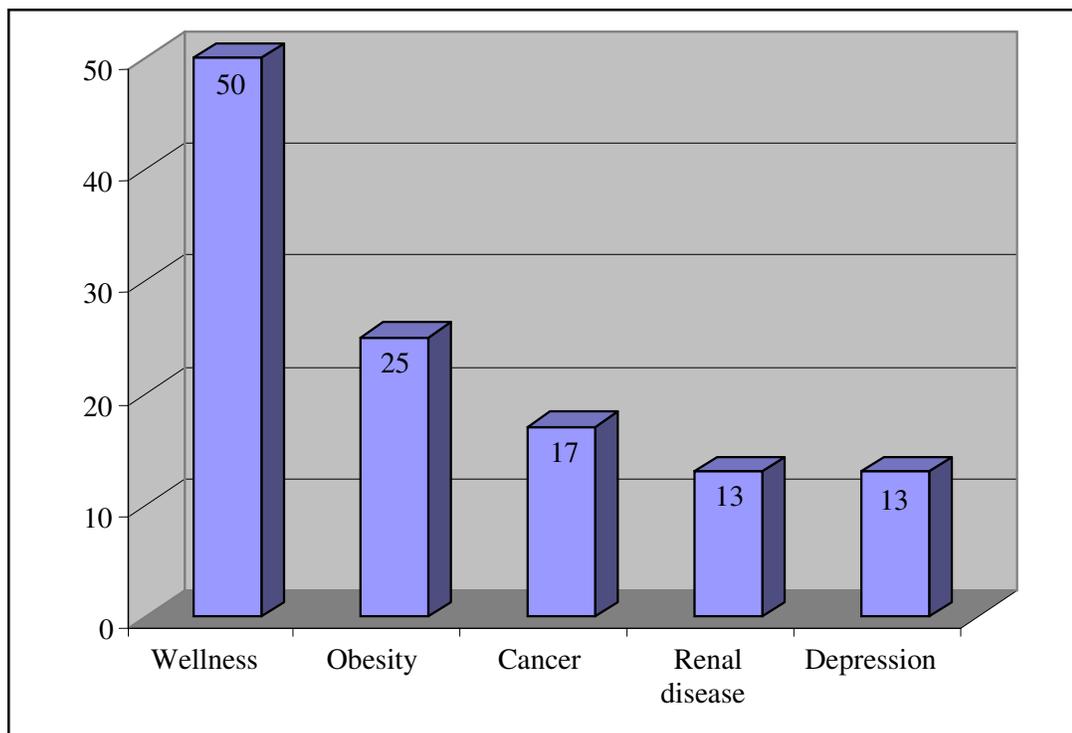
It is also reported by Matheson, et al. [48] that “health plans actually make the build-or-buy decision on a condition-by-condition basis. Harvard Pilgrim exemplifies this approach, by having internal programs for some conditions, such as asthma and diabetes, while contracting with one DMO for a cardiac program and another for rare diseases. Furthermore, some payers blend in-house resources and external services in the same program, for example, using in-house nurses in coordination with data analytics purchased from a vendor.” This shows that the transaction cost factors will affect each program in a health plan differently, leading to different organizational forms for each as the situation demands. They recommend that DMOs reduce transaction costs for health plans by “more effectively targeting and communicating to employer groups and health plans, and differentiating and marketing”. Employers and health plans are already requesting customized reporting on the outcomes of the DM programs, with greater detail in savings and health improvements which reduce uncertainty at the cost of higher transaction costs. They state that using efficient disease management programs, health plans and employers can leverage them strategically in order to build a competitive advantage. The most important element to make this possible is that “they should strive for excellence in the management of administrative and information technology costs”, both of which are components of transaction costs.

According to an article in Disease Management News [35], “Creating a successful disease management will require senior management commitment and dedicated resources”, and that “(disease management) programs are difficult because they require an unprecedented level of coordination, communication, and synthesis of information.” Both internal and contracted programs require time from senior management and commitment of capital and resources to be successful. In the case of contracted programs, it is seen that there needs to be close communication and information flow between many departments of both the health plan and the DMO to build a successful disease management program. The information systems department in both firms in particular, needs to have a close bond in order to develop the outcomes tracking and reporting functions. The medical directors of both firms also need to work together on the program protocols and integrating the program with the case management function [35]. Thus we can see that transaction costs are very prominent in the implementation of these programs and it is imperative that the organization choose a form as to minimize these costs.

We can conclude that all the major transaction cost factors which are asset specificity, frequency, and uncertainty will play a part in the final form adopted by a particular health plan for these programs.

### 2.3.4 Future of Disease Management

Disease management is also expanding worldwide, especially in Europe and Asia, due to its rapid growth in the U.S., particularly in the Medicare and Medicaid sectors. Australia has implemented many DM pilots recently, and Singapore has invested significantly in DM. Other countries implementing DM are Brazil and South Africa, whereas the United Kingdom and the Calgary health region in Canada are developing initiatives in DM [48]. Most of the DMOs and health plans are also looking to apply DM to additional areas such as obesity, cancer, and other cardiac conditions as they seek to achieve additional savings and meet employer demands for these programs, according to Matheson, et al., in 2006 [41]. Also, they are counting on increasing the number of people being covered by these programs, mainly by going deeper into the risk categories for each condition. Most health plan executives and decision makers view the DM industry to be in its growth phase [48]. Figure 2.5 shows the areas most likely for expansion and program development in the near future.



**Figure 2.5 Percentage of Health Plans and Respective Anticipated Areas of Expansion for DM Programs**

Source: Market data, BCG, Feb 2006. [48]

## 2.4 Selection Bias and the Heckman Two-Step Method

There are two forms of the selection bias problem. In the standard case of selection bias, information on the dependent variable for part of the respondents is missing. In the other version of the selection bias problem, information on the dependent variable is available for all respondents, but the distribution of respondents over categories of the independent variable we are interested in has taken place in a non-random manner.

Common to both forms of selection bias is that there is a selection process by which data is divided over two (or more) groups and that non-randomness in this process disturbs the estimation of other relationships which are of substantial interest. Thus, as described by Smits [64], there are two processes (which can be described with two equations, called selection equation and substantial equation) and these processes are related to each other. This relationship will be reflected in a non-zero correlation between the error terms of the equations. If such a correlation is present, we cannot estimate the substantial equation without taking the selection process into account. The Heckman two-step procedure can also be used to address both the forms of selection bias, and is taken from the classical papers of Heckman (1979, 1980).

This method was first derived by James Heckman in 1979 [30]. In this paper, the bias that results from the usage of non-randomly selected samples to estimate behavioral relationships as an ordinary specification error or omitted variable bias is discussed. The specification error framework is assumed to be the same as that specified by Griliches [25], Breen [7], and Theil [69]. He states that sample selection bias may arise for two reasons. First, there may be self selection by individuals or data units being studied. Second, sample selection decisions by researchers may lead to this bias. Using a computationally tractable technique, a simple consistent two stage estimator is considered that enables analysts to utilize simple regression methods to estimate behavioral functions using least squares method. The asymptotic distribution of the estimator is also derived.

In the first step of the Heckman procedure, the selection process which is responsible for selection bias problems is studied with the so-called selection model. For this purpose, generally a probit model is estimated (as the error term of this model is normally distributed, one of the assumptions underlying the Heckman model).

Next, the residuals of the selection equation are used to construct a selection bias control factor, which is called lambda. This factor is a summarizing measure which reflects the effects of all unmeasured characteristics which are related to the selection decision. Lambda is called the inverse mills ratio and is denoted as:  $f(z)/F(z)$ , where  $z$  is the estimated value from the probit equation and “ $f$ ” and “ $F$ ” denote the standard normal density and distribution functions, respectively. The value of this variable for each of the respondents is saved and used as an additional variable.

In the second step of the Heckman procedure, the main analysis is performed, in this case an ordinary least squares (OLS) regression analysis of the effects of sourcing decision on costs. In this substantial analysis we use the selection bias control factor calculated above as an additional independent variable. Because this factor reflects the effect of all the unmeasured characteristics which are related to the dependent variable of the initial model, the coefficients of this factor in the substantial analysis catches the part of the unmeasured characteristics related to the dependent variable in the secondary equation. Due to the presence of a control factor (lambda) in the analysis to compensate for the unmeasured characteristics of the dependent variable, which is also related to the dependent variables in the (initial) selection model, the predictors in the equation are freed from this effect and the regression analysis produces unbiased coefficients.

This method was first applied by Hanoch [26] in labor applications. In this industry, wages are observed only for those who actually work. However, one can infer from the decision to work and characteristics of the working laborers the reservation wage that most likely generated the pattern of observed employment and the observed wages at that time.

Heckman has applied his own methodology in his 1980 paper [31]. Here, he presents an empirically tractable model of the life cycle labor supply decisions of married women in an environment of perfect certainty. He integrates two distinct dimensions of life – time labor supply: annual hours worked and annual participation in the work force using his two – step approach, and using eight years of panel microdata from the Michigan panel Survey of Income Dynamics in order to estimate the model. Thus he extends the work done by Hanoch above, as that has stated only hours per week and hours per year as the two arbitrary dimensions.

He finds that labor supply is inversely related to life-time wealth measures, children affect life-time labor supply decisions, and that future values of variables determine current labor supply decisions. The usage of this methodology in this research has been detailed in the methodology section.

From the above literature review it seems very essential that further study of the sourcing decision of these programs be conducted. The proposed project builds on research in the study of factors affecting the outsourcing of disease management programs in a medical insurance organization. It focuses primarily on using transaction cost economics as a framework for better understanding the sourcing decisions and the internal organization costs and the external market costs that lead to this decision.

## **Chapter 3. Literature Summary and the Hypotheses**

### **3.1 Literature Summary and Application to Health Plans**

As the number of topics related to both transaction cost economics and disease management is very large, the researcher acknowledges that the literature review is not exhaustive, however the literature reviewed is sufficient to get a grasp on the key issues with which the research is concerned. These have been summarized here and used as a basis for the hypothesis detailed in this chapter.

We can infer from the literature review that all types of asset specificity, uncertainty and frequency affect the levels of transaction costs and hence affect the organizational form. Bounded rationality also places limits on the organization's ability to complete all activities internally or outsource completely and foresee and contract for all possible contingencies. Firms internalize their most important tasks and personnel to control quality and production, ensure access to scarce inputs, and have a better understanding of complex production/service techniques and technology. Based on the particular situation, firms should only integrate transactions that they can perform more effectively in-house than through contracting. This implies that if the total cost inclusive of the costs of selection, contract management, performance measurement, and dispute resolution are less than internal costs of providing the same good or service, then it must be outsourced, as the associated transaction costs are lower in that case.

According to transaction cost theory services formerly performed internally will tend to be outsourced if 1) the scale at which the service is performed efficiently increases relative to demand and 2) if the service becomes more standardized, less customer specific or more widely used.

Therefore, for an organization considering outsourcing there is not one clear answer regarding organizational form. It depends on the type of transaction and the specific conditions and factors that influence the organization and the industry.

### **3.2 Application of TCE Factors to Health Plans**

Transaction cost analysis has been applied to various manufacturing applications, which deals with continuous processing of a large quantity of material as they move from one processing station to the next, and construction industries, which involves the building of a single or unit at a fixed location, and the finished unit may or may not be made up of a small number of finished units. The various transaction cost factors and their effect on the organizational form in the healthcare sector can be hypothesized as follows:

In manufacturing, physical asset specificity is usually higher due to the high volume of production and the portability of the finished goods, compared to construction projects, where the final product is unique or produced in limited quantities, but the assets themselves are multipurpose and mobile. Disease management programs are mainly concerned with the monitoring of the individuals enrolled, which requires advanced software and computing power, and the provision of timely information to both the patient and the physician (which is done through various means of communication), hence, physical asset specificity is likely to be an important factor in the determination of the organizational form of a disease management program. We state hypothesis 1 such that integration of disease management programs becomes more likely as physical asset specificity increases.

Temporal specificity does not play a major role in the organizational form for manufacturing operations as it is of a high volume and continuous nature, whereas in the construction field a delay at one stage can reverberate through the entire project, and thus is more important in this application. The same can be said of the disease management, as it requires the timely dissemination of medical information both to the patient and to the physician. A delay in this regard could potentially lead to serious consequences to the afflicted person, and to the organization in the form of treatment costs, and hence this factor is likely to play an important role in the arrangement of the firm. We state hypothesis 2 such that integration of disease management programs becomes more likely as temporal specificity increases.

The factor site specificity can be explained as the distance between the interacting firms, Or Transportation and inventory costs specific to the transaction. Disease management programs are mainly concerned with the timely disposal of critical information to patients and physicians and coordination of medical services and tasks between the providers and the patients in order to provide best evidence care, and to make the patients active participants in their own care. Thus, these programs are not involved in delivering specific services or components at specific sites or individuals; hence, this factor is hypothesized to exert a very low influence on the outsourcing decision, and has thus not been considered in the empirical analysis. We state hypothesis 3 such that site specificity does not play an important role in the determination of the organizational form for disease management programs.

The factor dedicated assets can be defined as substantial, general-purpose investments specific to the transaction, and that need to be invested in for the proper completion of the transaction or service, or high-capacity equipment whose capacity is intended to be dedicated to a particular customer. In this context, dedicated assets may refer to capacity that is created to serve particular/specific customers, so that it would be difficult to find alternative customers, or an alternative use for the capacity created. Here, the effect of this factor will depend on both the disease being monitored and the size of the population enrolled. We state hypothesis 4 such that outsourcing/contracting of disease management programs will be more likely as the dedicated asset specificity rises.

The factor human asset specificity is generally not important in the manufacturing area due to the generalized and labor-intensive nature of the tasks involved. In the construction field, this factor may vary in importance, while generally it mirrors the construction field and the importance of this factor is low, however, there may be some construction applications (such as naval shipbuilding) may require specialized knowledge and skills, which increases the influence of this factor over the firm. Similarly, Human asset specificity will most likely exert a big influence over the organization structure as the experience, knowledge and skills needed for managing and running disease management programs are very specialized and specific. Usually, only experienced medical professionals (physicians and nurses) make up any given disease management team. We state hypothesis 5 such that outsourcing/contracting of disease management programs will be more likely as human asset specificity rises.

Uncertainty or Complexity also has a role to play in disease management programs. Disease management programs are generally very complex and require advanced knowledge of medical protocol, treatments and procedures. The symptoms and issues of the enrolled people will differ from person to person and this will lead to a degree of complexity much higher than that found in either the construction or the manufacturing areas, and is highly likely to play a major role in the form of the organization. We state hypothesis 6 such that programs with lower uncertainty and complexity will tend to be integrated while those which entail higher uncertainty and complexity will be contracted.

Disease management programs consist of many high-technology, medical knowledge intensive activities, unlike construction and manufacturing operations, where labor intensive, low-tech activities make up the bulk of the work. The tasks involved will also vary significantly on a patient to patient and also on a program-to-program basis. Thus, similarity in the disease management context is hypothesized to be low (between as well as within programs) and will likely play an important factor in determining organizational form. We state hypothesis 7 such that disease management programs similar to the ones already offered by the health plan are likely to be in-sourced, while those dissimilar to current programs will tend to be outsourced.

In this context, frequency refers to how often contact is made with the patients for interventions relating to their specific conditions. In other industries, it is seen that increased frequency leads to a greater probability of outsourcing or contracting to external vendors. The effect of this factor here is hypothesized to be similar, i.e. programs that require frequent contact will tend to be outsourced. We state hypothesis 8 such that the higher the frequency, the higher the chances of the disease management program being outsourced or contracted.

This study will focus on the factors outlined above and will involve collection of data based on the previously stated transaction cost factors as a means of constructing a probit regression model to study the effect of these factors on the form adopted by an organization for implementing disease management programs and to provide a dollar estimate of the costs borne by the organization.

The proposed study will focus on health management organizations who have implemented disease management programs both internally and through external vendors as a means to gather representative data based on the previously stated transaction cost factors as a means of

constructing a regression model to study the effect of these factors on organizational form, organization cost, and the role played by the major transaction cost factors in disease management programs through out the country and to provide a dollar estimate of the costs borne by the organization. Answers to the above questions can give a better insight into the issues of outsourcing from a health plan's perspective.

### **3.3 The Hypotheses**

Based on the understanding and the appreciation of this literature stated above, the researcher formed the following hypotheses to be tested:

- 1) Transaction cost factors yield influence over the organization form of disease management programs in managed care health plans.
- 2) The transaction cost factors exert their principal effect on the costs of internal organization, rather than external market costs.

The researcher's primary and secondary data collection is centered on testing these hypotheses. In order to test these hypotheses the researcher had to answer the four secondary questions outlined below and explained in the measures, instruments, and data sources section:

- 1) What is the nature of organization adopted for the disease management programs implemented by various health plans in the country?
- 2) What is the impact of transaction cost factors on integration decisions for disease management programs?
- 3) What are the implications for designing regression models for prediction of organization form and costs on the basis of transaction cost factors?
- 4) To analyze if selective organization leads to savings for the managed care organization or health insurance organization.

## Chapter 4. Methodology

### 4.1 Research Approach

The research approach and methodology used in this thesis will be based on quantitative data analysis collected by surveys and secondary data obtained from organizations in the health insurance industry such as managed care organizations, indemnity health plans, Medicare, etc.

The primary data and the secondary data will be collected by in-depth survey from the internal departments of willing health insurance organizations.

### 4.2 Research Method

Transaction cost analysis of organizational form maintains the hypothesis that the organization is so arranged as to minimize the cost of governing the transactions. The organization of the firm can be expressed as a binary variable, which is make or buy, that is, whether the component or service will be produced in-house or contracted/outsourced to an external vendor. There are two methodologies generally used for the measurement of transaction costs:

Direct Measurement: the first and most straightforward way of predicting the organizational form chosen would be by direct measurement and comparison of the costs, for example, if we denote the form chosen as  $F^*$ , a model of the choice between the two arrangements can be shown as:

$F^* = F_o$ , if  $C_o < C_m$ , and “ $F_o$ ” represents the integrated form for the task or service.  
=  $F_m$ , if  $C_o \geq C_m$ , and “ $F_m$ ” represents the outsourced form for the task or service.

where “ $C_o$ ” and “ $C_m$ ” represents the costs of internal production and market subcontracting respectively [47]. However, many costs such as inflexibility or need of litigation may not be addressed. Also, the most basic and fundamental problem in this approach is that organization

costs cannot be observed for the organizational forms not chosen. For example, if an organization chooses internal production and the associated costs are measured, the costs of organization for the alternative form i.e. outsourcing cannot be measured as that organizational form does not exist. Thus, direct cost observation is not a feasible method for the application of transaction cost analysis. In order to address this shortcoming, the following methodology has been adopted.

Reduced form analysis: in this methodology, the transaction costs in each possible organization form are related to observable features and then predictions of final adopted organizational form are made based on these features. Hence, the true costs of organization can be said to be:

$$C_o = AX + e, \quad (1)$$

$$C_m = BZ + u, \quad (2)$$

where  $X$  and  $Z$  are vectors of attributes (in this case, they are transaction cost factors) influencing the respective organizational costs,  $A$  and  $B$  are coefficient vectors and  $e$  and  $u$  are normally distributed random variables. Thus, the probability of observing organization form 'Fo' becomes:

$$F_o = \Pr(C_o < C_m) = \Pr(e - u < BZ - AX).$$

Thus, the comparison is now based on the signs and magnitudes of the coefficients  $A$  and  $B$ , and not on the direct costs ( $C_o$  and  $C_m$ ) themselves. However, if the variance of the difference between the random variables ( $e - u$ ) is not known, the coefficients of the above equations can only be identified up to a proportionality factor. Additionally, if  $X$  and  $Z$  share elements, only the differences between the vectors  $A$  and  $B$  can be identified [47]. As a result, it is not possible to deduce where the principal effect of the transaction cost factors lie, on internal or market costs. In order to obtain stronger tests of the theory, the method given below will be used in this analysis.

Two-stage analysis: as the name implies, this method consists of two stages. The first stage consists of the construction of a logistic or probit regression model as shown below. The logistic model takes the form:

$$\text{logit}(p_i) = \ln \left( \frac{p_i}{1 - p_i} \right) = \alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i},$$

$$i = 1, \dots, n,$$

Where “ $\alpha$ ” is the constant/intercept obtained from the model, “ $k$ ” represents the numbers of the independent variables “ $x$ ”, which each can have “ $i$ ” levels as shown above. “ $\beta$ ” is the parameter coefficients obtained for each of the independent variables from the model, and “ $p_i$ ” is the probability of the task or service being in-sourced. Thus, “ $p_i$ ” can be calculated as:

$$p_i = E(Y|X_i) = \Pr(Y_i = 1).$$

here,  $Y = 1$  for the in-house case, and

$Y = 0$  for the contracted case.

$$p_i = \Pr(Y_i = 1|X) = \frac{e^{\alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}}}{1 + e^{\alpha + \beta_1 x_{1,i} + \dots + \beta_k x_{k,i}}}.$$

The parameters  $\alpha, \beta_1, \dots, \beta_k$  are usually estimated by maximum likelihood.

The probit model assumes that:

$$\Pr(Y = 1|X = x) = \Phi(x'\beta),$$

where  $\Phi$  is the cumulative distribution function of the standard normal distribution,  $Y$  is the binary outcome variable, and  $X$  the vector of regressors. The parameters  $\beta$  are typically estimated by maximum likelihood. In this research we estimate a probit model in the first stage.

In this case, only the costs of organization for the form actually adopted are calculated. Thus, the model becomes:

$$C = C_o = AX + e, \text{ if } C_o < C_m,$$

$$C = C_m = \text{n.a.}, \text{ If } C_o \geq C_m.$$

In the second stage, switching regression techniques can be used to provide estimates of the internal organization costs. Estimation of the equations as a censored regression model will further reduce the need of large quantities of data. First, the inverse mills ratio (the ratio of the

probability density function over the cumulative distribution function of a distribution) is calculated as  $M = f(z)/F(z)$ , where “z” is the estimated probit values from the model, f = the probability density function and F = cumulative distribution function of a distribution.

The internal organization cost equation is constructed by regressing each transaction cost factor against our measured costs for in-sourcing. The equation for the internal organizational form is thus deduced by using the above equation and the inverse mills ratio.

The transaction cost methodology described above was applied to a specific application, which is, the outsourcing of disease management services by health plans and health management organizations (HMOs) to disease management organizations (DMOs). Whereas previous empirical research has dealt with manufacturing and construction applications, the process of disease management is quite different and removed from these, which in turn influences the circumstances that lead to opportunism and affects the nature of the organization and the associated costs.

### **4.3 Design and Study Participants**

A linear regression model featuring the decision to integrate as the dependent variable and the various transaction cost factors explained above as independent variables will be constructed. Health management organizations, including Medicare and Medicaid which engage in disease management plans and its outsourcing will be considered in this study. Initially, only those health management organizations situated and serving the population of Florida were considered. However, in order to obtain sufficient data, the sample size was expanded to include health management organizations from other states in the U.S. as well.

After the selection of the health organizations, data was collected for the construction of this model based on the disease management programs implemented for chronic diseases (see table 2.1). Pertinent data regarding any disease management program that was obtained was added to the construction of the model in addition to the basic five diseases. The collection of this data provides insight and better understanding of the effect that the considered factors of transaction cost have on the final organization form and allows estimation of the organization costs incurred with the current form and also under the other allowable alternative. Moreover, this model shows

the importance of internal organization costs in the outsourcing decision, which has previously only been applied to the construction and manufacturing fields, and never to the disease management field.

#### **4.4 Measures, Instruments and Data Sources**

The inquiry was guided by four research questions: 1) What is the nature of organization adopted for the disease management programs implemented by various health plans in the country; 2) what is the impact of transaction cost factors on integration decisions for disease management programs; 3) what are the implications for designing regression models for prediction of organization form and costs on the basis of transaction cost factors?; and 4) to analyze if selective organization leads to savings for the managed care organization or health insurance organization.

To determine the nature of organization in the various managed care health plans in the country that implement disease management programs (Question 1), an analysis of various health management organizations (including Medicare and Medicaid) was conducted. The number and types of disease management programs were noted and used to answer this question. The organization of the different plans (vertically integrated or subcontracted) were of particular interest. The purpose is to establish a frame of reference for the identification of the factors to be studied and included in the regression model.

To assess the impact of transaction cost factors on integration decisions for disease management programs (Question 2), all voluntary health management organizations were asked to complete a survey design based on previous surveys done by Monteverde and Teece (1982a) [51], Masten (1984) [45], and Anderson and Schmittlein [3]. These surveys were previously used to collect data from firms engaged in construction and manufacturing, such as the automotive and aerospace industry.

The survey covers a sample of tasks and services that can be integrated or outsourced by a health management organization while implementing a particular disease management program. The original survey has been extensively modified in order to adapt it to gather information on the disease management area, and it differs significantly from its original usage in the other industries. The original variables, the definitions, descriptive details and layout of the survey along with the modified version will be presented below. It is designed such that a team of

company officials such as the planning and implementation managers of the specific health plans and in some cases the higher management of the company can respond to each item on the survey based on their judgment. This data enables the construction of the linear regression model for the estimation of the decision to integrate services within the organization and the coefficients of the respective transaction cost factors in order to judge their importance.

Thirdly, to determine the implications for designing regression models for prediction of organization form and costs on the basis of transaction cost factors (Question 3), a summative analysis of the quantitative data was conducted, along with a comparison of the predictions made by our model with the actual organizational form, in order to determine the effectiveness of the constructed model, which also provides a more reliable picture of the performance of the model. In addition, data on organization costs were collected in order to estimate the cost of alternate arrangements.

Finally, to analyze if selective organization leads to savings for the managed care organization or health insurance organization (Question 4), we obtained the organization costs for the disease management programs that have been integrated into the organization, and also those that have been contracted to external vendors (DMOs). Thus, we can obtain estimates of organizational transaction costs for both the cases possible for these programs for evaluation and comparison of the costs incurred.

#### **4.5 Primary Data Collection**

As noted above, hypotheses regarding the effect of various transaction cost factors on the outsourcing decision for a disease management program have been put forth. To test these hypotheses, data was collected from health management organizations.

The independent variables corresponding to the hypotheses stated above are based on the respective transaction cost factors and are scored using a 5-point Likert scale and are explained in table 4.1 below. The specific questions asked in the survey have been detailed in appendix B at the end of the document.

**Table 4.1 Probit Model Variable Definitions and Descriptions**

Question	Variable	Definition
Q - 1	Disease class	The disease managed by the program.
Q - 2	Organization form	= 1, if the program was in - sourced, = 0, if the program was outsourced.
Q - 3-a, Q - 3-b, Q - 4, Q - 5	Measurement of transaction costs (Co)	This can be measured as the time spent in relation to the program X the average hourly management wage.
Q - 6	Temporal specificity	Ranking of the importance of timing of interventions, patient/program effectiveness checks, risk evaluations, etc.
Q - 7	Physical asset specificity	The degree to which the facilities and equipment is specific to the application.
Q - 8	Human asset specificity	The degree to which the knowledge, skills and experience of employees is specific to the application.
Q - 9	Dedicated asset specificity	A ranking of the degree of dedicated assets required for the program.
Q - 10	Complexity (proxy for uncertainty)	A ranking of the complexity of the tasks involved in the program.
Q - 11	Similarity	A variable that ranks a program according to the similarity of tasks with respect to the other programs run by the health plan, and how similar the required care is between the patient classes in the different programs.
Q - 12	Frequency	This variable ranks how often patient and physician contact is made by the program staff.
Q - 13	Uncertainty (proxy for uncertainty)	Provides a ranking of the difficulty in measuring the results, performance evaluation and effectiveness of the program.

In addition to these independent variables, data was collected on organization costs for the estimation of the structural cost equations as given in (1) and (2). The acquisition of this data has proved difficult and has varied based on the organization. For outside contracting particularly, this difficulty is exacerbated as outsourcing involves two parties and costs will be borne by both of them, necessitating the need for data to be collected from two sources. Also, contractual failures occur probabilistically over a period of time in the future, which leads to the data being collected being based on the views and expectations of the decision makers involved.

By contrast, costs of internal organization (planning, execution) occur in a single organization and in a more routine manner. Thus, these costs are easier to obtain or if actual measurement is not possible, reasonable proxies can be constructed. We thus concentrated on obtaining these costs, that is, the costs of internal organization for the processes and services actually done in house by the firm. Based on this, the costs of organization can be obtained by calculating the number of hours consumed by the decision makers for the planning, set up and execution of a service or process times the average hourly wage rate for the management involved, which was found to be \$60/hr after investigation into the industry and its associated wages.

In the designed survey, there are four questions pertaining to costs. As transaction costs are not usually measured and recorded, the questions ask for time estimates that are then converted to a monetary value. These questions are the questions 3a, 3b, 4 and 5. Question 3a is concerned with obtaining the time estimate for the administrative and facility planning tasks associated with an in-house program. If the program is to be integrated, there will need to have been substantial time and effort spent in order to fulfill the required administrative and startup tasks of starting a program from scratch. These costs will be unique to an in-house program and will not be present in the case of an outsourced program. Question 3b is used to measure the legal costs incurred while contracting a disease management program. When the decision is made to outsource, an appropriate contract needs to be drawn up between the two parties in order to define and put down the terms and conditions of the partnership. This will involve negotiations and bargaining between the parties involved which leads to an additional cost incurred for the contracted programs. This cost is unique to outsourced programs and will not occur in the case of programs built internally by the health plans.

The next question (question 4) is concerned with the measurement of search and information costs. This question is common to both cases (in-sourced and outsourced), as relevant information regarding the disease management program will need to be collected regardless of the form decided upon. In the integrated case, costs will be incurred in obtaining information about the tools and facilities required, the expertise needed, and the outcomes and benchmarks to be set. In the case of outsourced programs, it will involve the selection of a vendor that meets all the set requirements from all the choices available in the market.

Question 5 is used to record the supervisory costs that are an integral part of transaction costs. This question is again meant for both cases of organizational form, as in the integrated case, the effectiveness and outcomes of the implemented program will need to be monitored and changes will need to be made (if needed) to the internal staff and tools of the health plan. In the case of an outsourced program, time will be spent on monitoring the outcomes/results reported to the management by the external DMO, and changes or improvements may need to be worked out as needed based on the decisions of the health plan management. The dollar values for this question are annualized.

Thus, the in-sourced costs are calculated as follows:

$$\text{In-sourced costs} = (Q3a + Q4) * \text{working hours} * \text{average hourly management wage} + Q5 * \text{weeks/yr} * \text{Average hourly management wage}$$

The outsourced costs are calculated as follows:

$$\text{Outsourced costs} = Q3b + (Q4) * \text{working hours} * \text{average hourly management wage} + Q5 * \text{weeks/yr} * \text{Average hourly management wage}$$

#### **4.6 Data Analysis**

Upon collection of the data, analysis was conducted upon the gathered data in a two-step procedure.

In the first stage, a probit regression model was constructed for the estimation of the selection decision regarding whether the process will be done internally (integrated) or if the task will be

subcontracted (outsourced) to an outside service provider. The building of this model provided us with the coefficients of the various factors considered and explained in table 4.1, and their effect on the final organization decision and the form chosen. This stage of the analysis provided answers as to the importance of the various factors considered in the analysis and their influence on the outsourcing decision.

In the second stage, estimation of the structural equations of the model is carried out. First, we estimate the internal organization cost equation for each integrated program based on the costs obtained for each program from the organization and the values for each of the coefficients obtained from the first stage results. For this, using the sample of integrated services, we can estimate the coefficients for the internal organization cost equation by regressing our measure of internal organization costs against each of the independent variables. We also obtain the log specifications of the linear internal cost equation calculated, as the log specification will constrain the organization costs in the positive direction and also provide a better fit to the obtained data. Estimates of the transaction costs for the contracted disease management programs are also calculated as detailed in section 4.5 and compared with the in-house costs obtained.

Upon obtaining the cost estimations for both integrated and outsourced programs, the in-house cost equation is used to estimate the transaction costs for the integrated organizational form for each program, given its specific attributes. Thus, a comparison of the various organizational forms can be made. We obtain the predicted dollar value of the integrated transaction costs to compare with the costs for the organizational form actually adopted. The costs to the firm that would be incurred if all the tasks/services or programs were to be integrated can also be obtained and compared to find the costs or savings caused to the health plan under each organizational form.

## **Chapter 5. Numerical Results and Inference**

This chapter describes the application of transaction cost economics theory to the area of disease management programs in health plans. Section 5.1 reports the frequencies of the data set obtained from the electronic survey responses. In Section 5.2 the data analysis is performed. In Section 5.3 the data set is separated in to the training and validation sets. Section 5.4 onwards details the modeling, results and inference from the resulting models.

TCE analysis is applied to cost prediction for the integrated subset of the obtained data set. These health plans with integrated programs are selected for cost analysis because the costs associated with integrated programs are much more accurately measurable as compared to the outsourced subset.

### **5.1 Frequencies of Respondents and Corresponding DM Programs**

In order to collect relevant data for the construction of the required models for the analysis, an electronic survey was sent to the health plans that agreed to participate in the research over a period of two and a half months. The survey was sent to health plans across the nation in order to obtain the largest possible sample size for analysis, and only completed surveys with responses for all questions were included in the analysis. The frequencies of the resulting aggregated data set are as given below in tables 5.1, 5.2 and 5.3, which list the responding health plans, the programs implemented and the frequencies of the ranking questions based on the transaction cost factors respectively.

**Table 5.1 Responding Organizations and Number of Respective Responses**

Organization	Frequency	Cumulative Frequency
Ault International Medical Management, LLC	1	1
BCBSVT	6	7
Blue Cross Blue Shield of Florida	5	12
BlueCross BlueShield of Tennessee	5	17
CareGuide, Inc.	5	22
Contra Costa Health Plan	1	23
Direct Remedy Inc.	1	24
Florida Health Care Plans	4	28
Great-West healthcare	5	33
Health Alliance Plan	5	38
Health Integrated	1	39
Health Net Inc.	5	44
HealthPartners	5	49
Healthy Futures, Inc	1	50
Humana, Inc.	7	57
IMS Managed Care, Inc.	5	62
Independence Blue Cross	3	65
Interactive Performance Technologies LLC	2	67
Medica	4	71
Memphis Managed Care Corp	1	72
Miller & Huffman Outcome Architects, LLC	2	74
Mountain States Home Care	1	75
Parkland Community Health Plan	1	76
Partners HealthCare	1	77
QualChoice	5	82
Quality First Healthcare, Inc.	1	83
Solucia Inc	5	88
Utah Medicaid	1	89
VillageHealth Disease Management	2	91
WellPoint, Inc.	1	92
William Blair	1	93

**Table 5.2 Frequency of Corresponding DM Programs**

Disease	Frequency	Cumulative Frequency
Asthma	17	17
Chronic Obstructive Pulmonary Disease (COPD)	11	28
Congestive Heart Failure (CHF)	17	45
Coronary Artery Disease(CAD)	14	59
Diabetes	17	76
Low Back Pain	1	77
Other: End Stage Renal Disease	2	79
Other: 16 complex chronic conditions (e.g., Crohn's, Parkinson's, Multiple Sclerosis, Sickle Cell, etc.)	1	80
Other: CKD	1	81
Other: Cancer	2	83
Other: Complex Conditions	1	84
Other: High risk pregnancy	1	85
Other: Hypertension	2	87
Other: Integrated program for 5 conditions (asthma, diabetes, congestive heart failure, coronary artery disease, COPD)	1	88
Other: Our Synergy program covers 21 conditions	1	89
Other: Rare Diseases	1	90
Other: all chronic health conditions	1	91
Other: maternal child	1	92
Other: Pressure ulcers	1	93

**Table 5.3 Frequencies of the Organization Form**

FORM

FORM	Frequency	Cumulative Frequency
In-sourced/Integrated	40	40
Outsourced	53	93

DEP

DEP	Frequency	Cumulative Frequency
0	53	53
1	40	93

DEP is the binary variable corresponding to the organization form encountered for each DM program. Each health plan was requested to fill out one survey for each disease management program they offered, and the organizational form for each program was asked. The answer to this variable was converted to the dependent variable “DEP” for the purpose of modelling. DEP = 1, if the program is in - sourced/integrated by the health plan, and DEP = 0, if the program is outsourced. Table 5.4 reports the frequencies of each independent transaction cost factor.

**Table 5.4 Frequencies of the Eight TCE Factors**

<b>UNCERTAINTY</b>	Frequency	Cum. Freq.	<b>PHYSICAL</b>	Frequency	Cum. Freq.
1	5	5	1	18	18
2	12	17	2	4	22
3	33	50	3	41	63
4	34	84	4	15	78
5	9	93	5	15	93
<b>HUMAN</b>	Frequency	Cum. Freq.	<b>CAPITAL</b>	Frequency	Cum. Freq.
1	26	26	1	6	6
2	7	33	2	20	26
3	28	61	3	31	57
4	20	81	4	22	79
5	12	93	5	14	93
<b>COMPLEXITY</b>	Frequency	Cum. Freq.	<b>SIMILARITY</b>	Frequency	Cum. Freq.
1	4	4	1	14	14
2	4	8	2	11	25
3	12	20	3	17	42
4	56	76	4	18	60
5	17	93	5	33	93
<b>FREQUENCY</b>	Frequency	Cum. Freq.	<b>TEMPORAL</b>	Frequency	Cum. Freq.
2	7	7	2	1	1
3	52	59	3	30	31
4	16	75	4	29	60
5	18	93	5	33	93

## 5.2 Data Analysis

In order to ensure the validity of the data, the statistics of the data and the correlations between the eight independent transaction cost variables each and also the correlation of each independent factor with the dependent variable needs to be checked. This is done by checking the Pearson correlation coefficient between all nine variables involved in the construction of the model. The results are as given below.

Pearson Correlation Coefficients – This statistic measures the strength and direction of the linear relationship between the two variables. The correlation coefficient can range from -1 to +1, with -1 indicating a perfect negative correlation, +1 indicating a perfect positive correlation, and 0 indicating no correlation at all. (A variable will always have a correlation coefficient of 1 with itself.)

N = 93 - This indicates that 93 observations were used in the correlation of each pair of variables.

Prob > |r| under H<sub>0</sub>: Rho=0 - This is the p-value and indicates the probability of observing this correlation coefficient or one that is more extreme under the null hypothesis (H<sub>0</sub>) that the correlation (Rho) is 0. The section is constructed in a way so that the top number is the correlation coefficient and the bottom number is the p-value.

The results of the procedure are presented in tables 5.5 and 5.6 below. Table 5.5 shows the means and statistics for all nine variables involved, while table 5.6 reports the results of the correlation procedure carried out where each variable is checked for correlation with itself and the eight others included in the analysis.

**Table 5.5 Means and Statistics for all Modeling Variables**

The CORR Procedure

9 Variables: DEP TEMPORAL PHYSICAL HUMAN CAPITAL  
 COMPLEXITY SIMILARITY FREQUENCY UNCERTAINTY

Variable	N	Mean	Std Dev	Sum	Minimum	Maximum	Label
DEP	93	0.43011	0.49777	40	0	1	DEP
TEMPORAL	93	4.01075	0.85331	373	2	5	TEMPORAL
PHYSICAL	93	3.05376	1.27999	284	1	5	PHYSICAL
HUMAN	93	2.83871	1.38541	264	1	5	HUMAN
CAPITAL	93	3.19355	1.135	297	1	5	CAPITAL
COMPLEXITY	93	3.83871	0.9242	357	1	5	COMPLEXITY
SIMILARITY	93	3.48387	1.45672	324	1	5	SIMILARITY
FREQUENCY	93	3.48387	0.89215	324	2	5	FREQUENCY
UNCERTAINTY	93	3.32258	1.00175	309	1	5	UNCERTAINTY

**Table 5.6 Correlations for all Modeling Variables**

Pearson Correlation Coefficients, N = 93

Prob > |r| under H0: Rho=0

	DEP	TEMPORAL	PHYSICAL	HUMAN	CAPITAL
DEP	1	0.09135	0.11685	-0.00864	0.37051
	-	0.3838	0.2647	0.9345	0.0003
TEMPORAL	0.09135	1	0.23831	0.24054	0.05394
	0.3838	-	0.0214	0.0202	0.6076
PHYSICAL	0.11685	0.23831	1	0.086308	0.09002
	0.2647	0.0214	-	<.0001	0.3908

	COMPLEXITY	SIMILARITY	FREQUENCY	UNCERTAINTY
DEP	-0.27286	-0.1852	-0.10659	0.0675
	0.0081	0.0755	0.3092	0.5203
TEMPORAL	0.15383	-0.4415	0.19298	-0.10583
	0.141	<.0001	0.0638	0.3127
PHYSICAL	0.09929	-0.0957	-0.01351	0.09653
	0.3437	0.3614	0.8977	0.3573

	DEP	TEMPORAL	PHYSICAL	HUMAN	CAPITAL
HUMAN	-0.00864	0.24054	0.086308	1	0.02007
	0.9345	0.0202	<.0001	-	0.8486
CAPITAL	0.37051	0.05394	0.09002	0.02007	1
	0.0003	0.6076	0.3908	0.8486	-
COMPLEXITY	-0.27286	0.15383	0.09929	0.0304	-0.05281
	0.0081	0.141	0.3437	0.7724	0.6151
SIMILARITY	-0.1852	-0.4415	-0.09572	-0.21943	-0.02439
	0.0755	<.0001	0.3614	0.0346	0.8165
FREQUENCY	-0.10659	0.19298	-0.01351	0.06383	-0.05056
	0.3092	0.0638	0.8977	0.5433	0.6303
UNCERTAINTY	0.0675	-0.1058	0.09653	-0.0091	0.25041
	0.5203	0.3127	0.3573	0.931	0.0155

**Table 5.6 (Continued)**

	COMPLEXITY	SIMILARITY	FREQUENCY	UNCERTAINTY
HUMAN	0.0304	-0.2194	0.06383	-0.0091
	0.7724	0.0346	0.5433	0.931
CAPITAL	-0.05281	-0.0244	-0.05056	0.25041
	0.6151	0.8165	0.6303	0.0155
COMPLEXITY	1	-0.0948	0.20115	0.35032
	-	0.366	0.0532	0.0006
SIMILARITY	-0.0948	1	-0.35775	-0.01129
	0.366	-	0.0004	0.9144
FREQUENCY	0.20115	-0.3578	1	-0.26168
	0.0532	0.0004	-	0.0113
UNCERTAINTY	0.35032	-0.0113	-0.26168	1
	0.0006	0.9144	0.0113	-

If an independent variable is heavily correlated with another independent variable, one of them can be removed as both produce the same effect in the model and it is not necessary that both variables be included in the model. If an independent variable is heavily correlated with the dependent variable of the model, then the effect of the independent factor can be explained by the nature of the correlation between the two variables. However, from the above results we see that none of the independent variables are correlated strongly with each other and neither is any independent variables strongly correlated with the dependent variable (DEP). Hence all 8 of them can be included in the analysis.

### **5.3 Creating the Training and Validation Sets for the First Stage Selection Model**

The total sample size consists of 93 data points. The responses cover a large number of the disease management programs offered by health plans. The responses also show that there was no clear consensus as to which organizational form is better for these programs, as already evidenced from the literature review. For the sample obtained for the purpose of this research, it is seen that outsourced programs (N = 53) slightly outnumber the integrated cases (N = 40).

This data set was randomly split into two groups: the training set on which the selection model was built and the inference was deduced, consisting of 80 observations and the validation set, which was used to test the model and determine its accuracy, containing 13 observations. The frequencies for the two sets are as given below in tables 5.7 and 5.8.

**Table 5.7 Frequencies for the Training Set**

The FREQ Procedure

DEP

DEP	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	47	58.75	47	58.75
1	33	41.25	80	100

**Table 5.8 Frequencies for the Validation Set**

ACTUAL_FORM	Frequency	Percent	Cumulative Frequency	Cumulative Percent
0	6	46.15	6	46.15
1	7	53.85	13	100

## 5.4 Training Stage

### 5.4.1 Step 1. Creating the First Stage Selection Model

The first step is to create the selection model. The results of this model will provide answers to the following questions:

- 1) Do transaction cost factors affect and influence the sourcing decision for disease management programs?
- 2) If so, which factors play the most important role in determining organizational form and what is their effect?

The combined set of responses needs to be split into two for the purposes of modeling. The first is the training set (with 80 randomly selected observations) on which the model is built and the coefficients and significance of the independent factors is noted. The second set is the testing set, which has the remaining 13 observations, where the model accuracy will be tested by comparing the predicted organizational form with the actual, which is known in our case. The training set created from the total sample is used to create the first stage probit selection model.

To compute the first stage selection model, the command “proc probit” can be used in SAS. This procedure does not allow for the classification table to be obtained, however, which is very helpful for checking the model accuracy in this case. As an alternative, a the “proc logistic” command is used along with the “link = probit” command. This command estimates a probit model based on the given data, while also allowing for the probabilities for each observation and the classification table to be constructed.

In this research based on the effect of transaction cost factors on organizational form of disease management programs, the selection model contains the eight independent transaction cost variables. The dependent variable DEP is an indicator variable with value 1 for integrated programs and a value 0 for outsourced programs. The SAS commands are as follows:

```
proc logistic data=TRAINING descending;  
model DEP = temporal physical human capital complexity  
similarity frequency uncertainty/LINK=PROBIT ctable  
pprob=(0.05 to 1 by 0.05);  
output out=prob XBETA= g predicted=phat;  
TITLE 'FIRST STAGE SELECTION MODEL';  
run;
```

Where, “training” is the data set containing the 80 data points, the command “XBETA” gives us the probit scores generated for each of the observations. The command “predicted” provides us with the probabilities for each of the observations recorded. In the output of this analysis, we find the estimates of the parameters. On the basis of these parameters, for each observation the predicted probit score is also obtained, which is stored in the variable “g”.

The command “predicted” gives us the probability values calculated for each of the observations in the training set. The results are as given below. Table 5.9 gives the model fit statistics and the significance of the probit model for to the data using the statistical parameters given below.

**Table 5.9 Full Model Response, Fit Statistics and Null Hypothesis for Training Stage**

FIRST STAGE SELECTION MODEL

The LOGISTIC Procedure

Model Information

Data Set                    WORK.TRAINING  
 Response Variable        DEP                    DEP  
 Number of Response Levels   2  
 Model                      binary probit  
 Optimization Technique   Fisher's scoring

Number of Observations Read    80  
 Number of Observations Used    80

Response Profile

Ordered Value	DEP	Total Frequency
1	1	33
2	0	47

Probability modeled is DEP=1.

**Table 5.9 (Continued)**

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	110.441	98.614
SC	112.823	120.052
-2 Log L	108.441	80.614

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.8271	8	0.0005
Score	23.6146	8	0.0027
Wald	18.0993	8	0.0205

An explanation of the terms and results is given below:

- 1) Data Set - The data set used in this procedure.
- 2) Response Variable - The response variable in the logistic regression.
- 3) Number of Response Levels - The number of levels our response variable has. Here we have DEP = 0 and DEP = 1.
- 4) Model - The type of regression model that was fit to our data.
- 5) Optimization Technique - This refers to the iterative method of estimating the regression parameters. In SAS, the default is method is Fisher's scoring method

- 6) Number of Observations Read and Number of Observations Used - The number of observations read and the number of observation used in the analysis. The Number of Observations Used may be less than the Number of Observations Read if there are missing values for any variables in the equation. By default, SAS does a list wise deletion of incomplete cases. We see that all 80 observations of the training set have been used for the construction of the model, and none have been deleted.
- 7) Ordered Value - Ordered value refers to how SAS orders/models the levels of the dependent variable. When the descending option is specified in the procedure statement, SAS treats the levels of DEP in a descending order (high to low). By default SAS models the 0's (the outsourced cases). The descending option is necessary so that SAS models the 1's, that is, the integrated cases.
- 8) Total Frequency - The frequency distribution of the response variable. Our response variable has 33 observations with a DEP = 1 and 47 with DEP = 0.
- 9) Probability modeled is DEP = 1 - This is a note informing which level of the response variable we are modeling.
- 10) Model Convergence Status - this describes whether the maximum-likelihood algorithm has converged or not, and what kind of convergence criterion is used to assess convergence. The default criterion is the relative gradient convergence criterion (GCONV), and the default precision is  $10^{-8}$ .
- 11) Criterion – this lists various measurements used to assess the model fit, which consists of the following:
  - 1) AIC - The Akaike Information Criterion. It is calculated as  $AIC = -2 \log L + 2((k-1) + s)$ , where k is the number of levels of the dependent variable and s is the number of predictors in the model. AIC is used for the comparison of models from different samples or non-nested models. The model with the smallest AIC is considered the best.
  - 2) SC - This is the Schwarz Criterion. It is defined as  $-2 \log L + ((k-1) + s) \log(\sum f_i)$ , where  $f_i$ 's are the frequency values of the  $i^{\text{th}}$  observation, and k and s are as defined previously. Like AIC, SC penalizes for the number of predictors in the model and the smallest SC is most desirable.
  - 3) Akaike Information Criterion (AIC) and Schwarz Criterion (SC) are deviants of negative two times the Log-Likelihood ( $-2 \log L$ ). AIC and SC penalize the log-likelihood by the number of predictors in the model.

- 4)  $-2 \text{ Log } L$  - This is negative two times the log-likelihood. The  $-2 \text{ Log } L$  is used in hypothesis tests for nested models.
- 12) Intercept Only - This column refers to the respective criterion statistics with no predictors in the model, i.e., just the response variable.
- 13) Intercept and Covariates - This column corresponds to the respective criterion statistics for the fitted model. A fitted model includes all independent variables and the intercept. We can compare the values in this column with the criteria corresponding Intercept Only value to assess model fit/significance.
- 14) Test - These are three asymptotically equivalent Chi-Square tests. They test against the null hypothesis that at least one of the predictors' regression coefficient is not equal to zero in the model.
- 15) Likelihood Ratio - The Likelihood Ratio (LR) Chi-Square tests that at least one of the predictors' regression coefficient is not equal to zero in the model. The LR Chi-Square statistic can be calculated by  $-2 \text{ Log } L(\text{null model}) - 2 \text{ Log } L(\text{fitted model})$ , where  $L(\text{null model})$  refers to the Intercept Only model and  $L(\text{fitted model})$  refers to the Intercept and Covariates model.
- 16) Score - The Score Chi-Square tests that at least one of the predictors' regression coefficients is not equal to zero in the model.
- 17) Wald - The Wald Chi-Square tests that at least one of the predictors' regression coefficients is not equal to zero in the model.
- 18) Chi-Square, DF and  $\text{Pr} > \text{ChiSq}$  - The Chi-Square test statistic, Degrees of Freedom (DF) and associated p-value ( $\text{PR} > \text{ChiSq}$ ) corresponding to the specific test that all of the predictors are simultaneously equal to zero. The null hypothesis is that all of the regression coefficients in the model are equal to zero. The DF defines the distribution of the Chi-Square test statistics and is defined by the number of predictors in the model.  $\text{PR} > \text{ChiSq}$  is compared to a specified alpha level (willingness to accept a type I error), which is often set at 0.05 or 0.01.

#### **5.4.1 Step 2. Evaluate Results of the Training Stage**

From the above results, we see that the model is a good fit for the data obtained via the survey. It is now necessary to obtain the coefficients for each independent transaction cost factor in order to gauge their effect on the final organizational form chosen by the health plan for. The statistics and results obtained in this step are explained below:

- 1) Parameter – this column lists the predictor variables in the model and the intercept.
- 2) DF - This column gives the degrees of freedom corresponding to the Parameter. Each Parameter estimated in the model requires one DF and defines the Chi-Square distribution to test whether the individual regression coefficient is zero, given the other variables in the model.
- 3) Estimate - The binary probit regression estimates for the Parameters in the model.
- 4) Intercept - The probit regression estimate when all variables in the model are evaluated at zero.
- 5) Standard Error - The standard errors of the individual regression coefficients.
- 6) Wald Chi-Square and Pr > ChiSq - The test statistics and p-values, respectively, testing the null hypothesis that an individual predictor's regression coefficient is zero, given the other predictor variables are in the model.
- 7) Percent Concordant - A pair of observations with different observed responses is said to be concordant if the observation with the lower ordered response value (DEP = 0) has a lower predicted mean score than the observation with the higher ordered response value (DEP = 1).
- 8) Percent Discordant - If an observation with the lower ordered response value has a higher predicted mean score than the observation with a higher ordered response value, then the pair is discordant.
- 9) Percent Tied - A pair of observations with different responses is neither concordant nor discordant, and is termed a tied pair.
- 10) Pairs - The total number of distinct pairs with one case having a positive response (DEP = 1) and the other having a negative response (DEP = 0). The total number ways the 93 observations can be paired up (excluding be matched up with themselves) is  $93(92)/2 = 4278$ .
- 11) Somers' D - Somer's D is used to determine the strength and direction of relation between pairs of variables. Its values range from -1.0 (all pairs disagree) to 1.0 (all pairs agree). It is defined as  $(n_c - n_d)/t$  where  $n_c$  is the number of pairs that are concordant,  $n_d$  the number of pairs that are discordant, and  $t$  is the number of total number of pairs with different responses.
- 12) Gamma - The Goodman-Kruskal Gamma method does not penalize for ties on either variable. Its values range from -1.0 (no association) to 1.0 (full association). Because it

does not penalize for ties, its value will generally be greater than the values for Somer's D.

- 13) Tau-a - Kendall's Tau-a is a modification of Somer's D that takes into the account the difference between the number of possible paired observations and the number of paired observations with a different response. It is defined to be the ratio of the difference between the number of concordant pairs and the number of discordant pairs to the number of possible pairs  $(2(n_c - n_d)/(N(N-1)))$ . Tau-a is usually smaller than Somer's D since there are many paired observations with the same response.
- 14) c - c ranges from 0.5 to 1, where 0.5 corresponds to the model randomly predicting the response, and a 1 corresponds to the model perfectly predicting the response.

Through this step we obtain the parameter estimates for each transaction cost factor, as presented in table 5.10.

**Table 5.10 Analysis of Parameter Coefficients for the Training Stage**

FIRST STAGE SELECTION MODEL  
 The LOGISTIC Procedure  
 Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.6295	1.6646	0.9583	0.3276
TEMPORAL	1	0.0476	0.2179	0.0478	0.827
PHYSICAL	1	0.6327	0.2587	5.9824	0.0144
HUMAN	1	-0.5188	0.2496	4.3196	0.0377
CAPITAL	1	0.5214	0.1813	8.27	0.004
COMPLEXITY	1	-0.6031	0.2401	6.3088	0.012
SIMILARITY	1	-0.2788	0.1416	3.873	0.0491
FREQUENCY	1	-0.1736	0.2283	0.5785	0.4469
UNCERTAINTY	1	-0.067	0.2079	0.104	0.7471

Association of Predicted Probabilities and Observed Responses

Percent Concordant	80.1	Somers' D	0.603
Percent Discordant	19.7	Gamma	0.605
Percent Tied	0.2	Tau-a	0.296
Pairs	1551	c	0.802

### **5.4.1 Step 3. Use the Classification Table to Determine Optimal Cut-Off Point**

In order to maximize the effectiveness of the model in classifying events and non events for the validation set, we need to select a cut-off point for the predicted probabilities, one below which the program will be classified as outsourced, and above which the program will be classified as in-sourced. This comparison table was created using the “pprob” option in the modeling syntax available in the SAS software. Each event can be classified as a true positive or a false positive according to the following definitions:

- 1) Event: if the organizational form of a given DM program is predicted as in-sourced/integrated then it is termed as an event.
- 2) Non–event: if the organization form of a given DM program is predicted as outsourced then it is termed as a non-event.
- 3) False POS (False positive): if an event identified by the model is not an integrated program it constitutes a false positive.
- 4) False NEG (False negatives): if the non-event identified by the model as outsourced program is actually an event (in-sourced organizational form) it is termed as a false negative.

The results are as shown in table 5.11.

**Table 5.11 Classification Table for Training Set**

Prob Level	Correct		Incorrect		Percentages				
	Event	Non - Event	Event	Non - Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.05	33	3	44	0	45	100	6.4	57.1	0
0.1	32	11	36	1	53.8	97	23.4	52.9	8.3
0.15	32	11	36	1	53.8	97	23.4	52.9	8.3
0.2	25	16	31	8	51.3	75.8	34	55.4	33.3
0.25	24	19	28	9	53.8	72.7	40.4	53.8	32.1
0.3	23	21	26	10	55	69.7	44.7	53.1	32.3
0.35	23	25	22	10	60	69.7	53.2	48.9	28.6
0.4	23	30	17	10	66.3	69.7	63.8	42.5	25
0.45	23	38	9	10	76.3	69.7	80.9	28.1	20.8
0.5	20	39	8	13	73.8	60.6	83	28.6	25
0.55	17	44	3	16	76.3	51.5	93.6	15.0	26.7
0.6	15	45	2	18	75	45.5	95.7	11.8	28.6
0.65	14	45	2	19	73.8	42.4	95.7	12.5	29.7
0.7	14	45	2	19	73.8	42.4	95.7	12.5	29.7
0.75	8	45	2	25	66.3	24.2	95.7	20	35.7
0.8	7	45	2	26	65	21.2	95.7	22.2	36.6
0.85	7	46	1	26	66.3	21.2	97.9	12.5	36.1
0.9	7	47	0	26	67.5	21.2	100	0	35.6
0.95	4	47	0	29	63.8	12.1	100	0	38.2
1	0	47	0	33	58.8	0	100	.	41.3

From the above we see that we have the best prediction and highest value of correct classifications (and corresponding lowest number of false positives and negatives) occurs at the 0.55 probability level, hence that value is selected as the cut-off point to be used for the validation set.

## 5.5 Validation Stage

### 5.5.1 Step 1. Embed the Validation Set Into the Training Set

In this step we combine the training and validation data sets into one, but we leave the dependent variable information as unknown for the validation set.

When the model is run, the model is built again on the basis of the training set. However, the predicted probabilities for the validation set are also calculated and displayed. The SAS commands are as follows:

```
proc logistic data=COMBINED descending;  
model DEP = temporal physical human capital complexity  
similarity frequency uncertainty/LINK=PROBIT ;  
output out=prob2 XBETA= g2 predicted=phat2;  
TITLE 'FIRST STAGE SELECTION MODEL FOR COMBINED DATA SET';  
run;
```

Where, “combined” is the data set containing all the 93 data points. Using the cut-off point described above, one can then classify them as in-sourced or outsourced, and a comparison with the actual form (known in our case) can be made, if this information is stored in another variable (for this analysis, the actual form is stored in the variable “actual\_form”. The initial output for this step again details the number of observations used and the model fit statistics with the terms as explained in section 5.4.1 (step 1). The results are detailed in table 5.12, and show that out of the total 93 observations used, only the original 80 are used to construct the model, whereas the newly added training observations are not used as they have the dependent variable (DEP) as missing. However, predicted probabilities are still calculated for the testing set as well as this set contains all the independent variable values for each observation. Thus, a comparison of the predicted and actual form can be made in the later stages.

**Table 5.12 Full Model Response, Fit Statistics and Null Hypothesis for Testing Phase**

FIRST STAGE SELECTION MODEL FOR COMBINED DATA SET

The LOGISTIC Procedure

Model Information

Data Set                    WORK.COMBINED  
Response Variable            DEP                    DEP  
Number of Response Levels    2  
Model                        binary probit  
Optimization Technique      Fisher's scoring

Number of Observations Read    93

Number of Observations Used    80

Response Profile

Ordered Value	DEP	Total Frequency
1	1	33
2	0	47

Probability modeled is DEP=1.

**Table 5.12 (Continued)**

NOTE: 13 observations were deleted due to missing values for the response or explanatory variables.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	110.441	98.614
SC	112.823	120.052
-2 Log L	108.441	80.614

FIRST STAGE SELECTION MODEL FOR COMBINED DATA SET

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.8271	8	0.0005
Score	23.6146	8	0.0027
Wald	18.0993	8	0.0205

### 5.5.1 Step 2. Results of the First Stage Selection Model With Combined Data Set

As in the earlier case with the training set, in this step we obtain the parameter estimates for the transaction cost variables included in the modeling. The terms and statistics are the same as those explained in section 5.4.1 step 2. Table 5.13 reports the parameter coefficients for the independent factors.

**Table 5.13 Analysis of Parameter Coefficients for Testing Stage**

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	1.6295	1.6646	0.9583	0.3276
TEMPORAL	1	0.0476	0.2179	0.0478	0.827
PHYSICAL	1	0.6327	0.2587	5.9824	0.0144
HUMAN	1	-0.5188	0.2496	4.3196	0.0377
CAPITAL	1	0.5214	0.1813	8.27	0.004
COMPLEXITY	1	-0.6031	0.2401	6.3088	0.012
SIMILARITY	1	-0.2788	0.1416	3.873	0.0491
FREQUENCY	1	-0.1736	0.2283	0.5785	0.4469
UNCERTAINTY	1	-0.067	0.2079	0.104	0.7471

Association of Predicted Probabilities and Observed Responses

Percent Concordant	80.1	Somers' D	0.603
Percent Discordant	19.7	Gamma	0.605
Percent Tied	0.2	Tau-a	0.296
Pairs	1551	c	0.802

### 5.5.2 Classification of the Training Set

As we have selected the optimum cut-off point for classifying the observations in the model, we can classify the output according to the predicted probabilities for each data point in the training set to check the accuracy of the model on the training set. From table 5.11, it is seen that selecting a cut-off point of 55% as the demarcation between integrated and outsourced organizational form gives us the most correct classifications and the least number of false positives and false negatives. This cut-off point is now used on the training and validation set. Those observations from both sets that have a predicted value of less than 0.55 are classified as outsourced and the one's that have a predicted value of 0.55 or higher are classified as having an integrated organizational form. The results for both the training and the validation set can now be checked for accuracy. The results for the training set are presented in table 5.14. The first column denotes the predicted form, classified on the basis of the cut-off point, whereas the first row denotes the dependent variable (DEP), which is the actual organizational form for the disease management programs. The diagonal elements represent the frequency of the correct classifications, while the non-diagonal elements are the number of observations that have been erroneously classified.

**Table 5.14 Prediction Accuracy for the Training Set**

The FREQ Procedure  
Table of PRED\_FORM by DEP

Frequency	0	1	Total
0	45	13	58
1	2	20	22
Total	47	33	80

Frequency Missing = 13

Hence, prediction accuracy =  $(45+20)/80 = 81.25 \%$

### 5.5.3 Classification of the Validation Set

We follow the same procedure as in section 5.5.2 in order to check the classification of the validation set, which is the true measure of model effectiveness. The outcome is reported in table 5.15. Based on the classification process detailed earlier, it is seen that 76.92% of the observations in the testing/validation set are correctly classified. Thus, the model is accurate in the prediction of organizational form of disease management programs in health plans.

**Table 5.15 Prediction Accuracy for the Validation Set**

The FREQ Procedure

Table of PRED\_FORM by ACTUAL\_FORM

Frequency	0	1	Total
0	5	2	7
1	1	5	6
Total	6	7	13

Frequency Missing = 80

Hence, prediction accuracy =  $(5+5)/13 = 76.92\%$

### 5.5.4 Inference for the First Stage Selection Model

In table 5.13, results for the probit estimation of the decision to integrate the disease management production using the proxies for the seven transaction cost factors are shown. Of these factors, the coefficient for temporal specificity is positive as expected, indicating that the program is more likely to be integrated the more critical the scheduling of a task or service is to the program.

The coefficient for temporal specificity (TEMPORAL) is positive, meaning that as the importance of scheduling of the various tasks in the program rises, the more likely the program is to be integrated and not outsourced. The insignificance of the factor may be due to the fact that scheduling does not play as vital a role in DM programs as it does in other industries such as the automotive and ship building industries, and does not have a reverberating or domino effect on the rest of the program if the completion of a particular task or service is not according to schedule. Thus, according to the results, while our hypothesis detailed in section is correct to the extent of the effect of this factor on the organization form, the degree of importance that was hypothesized was overstated. Thus, hypothesis 2 from section 3.2 is only partially satisfied.

The coefficient for physical asset specificity (PHYSICAL) is also positive and significant, meaning that as the more specific the tools and assets used in the program; the more likely the program is to be integrated and not contracted. Results show that hypothesis 1 is completely satisfied and that the effect of physical asset specificity is as stated in section 3.2.

Another factor that demonstrates a similar effect is “CAPITAL”. The coefficient for dedicated asset specificity (CAPITAL) is also positive and significant, which supports the hypothesis that integration is more likely for programs that require specific investments that are unusable for any other purposes.

It is seen that hypothesis 4 is not satisfied and that the effect of dedicated asset specificity is the converse of what was detailed in section 3.2. It’s significance is correctly stated.

The coefficient for Human asset specificity (HUMAN) is negative and significant, which supports the hypothesis that contracting/outsourcing is more likely for programs needing specific skill sets and experience from the employees. Thus, hypothesis 5 is proved correct as the results from the above model match the effect detailed for this factor in section 3.2.

The coefficient for Uncertainty (COMPLEXITY) is negative and significant, which again supports the hypothesis detailed in section 3.2 that contracting/outsourcing is more likely for programs where the outcome reporting and performance measurement may be more difficult for the health plan, which may lead to higher transaction costs if such a program were integrated. Specifically, increases in complexity make it less likely that the program will be integrated within the firm. Hypothesis 6 is thus satisfied and results support our claim made in the earlier section.

The coefficient for similarity (SIMILARITY) is negative and significant, meaning that DM programs which are dissimilar to the ones the health plan may be involved in have a higher chance of integration than ones which may be similar to the ones already offered by the organization. The significance of this factor is correctly predicted, however, the effect on organizational form on DM programs is converse of that noted, leading to a partial validation of hypothesis 7.

The coefficient for frequency (FREQUENCY) is again negative but insignificant, meaning that as the frequency of contact required with the patents enrolled in the program rises, the probability of contracting the DM program to an external DMO rises. The insignificance of the factor may be explained by the fact that frequency within a program may vary significantly based on the individual characteristics of the patient and hence may not be a major factor in the sourcing decision for health plans. The effect agrees with our hypothesis stated in section in 3.2. Hypothesis 8 is thus partially satisfied. While the effect is concurrent with the stated hypothesis, the degree of effect exerted by this factor on the sourcing decision is not as high as was postulated.

Finally, the coefficient for uncertainty (UNCERTAINTY) is again negative but insignificant. This factor has already been covered by the independent variable “COMPLEXITY” which is a proxy for uncertainty. We see that the effect is similar to the effect of “complexity”, i.e. as the uncertainty of effectiveness and outcomes in a DM program rises, health plans tend to contract rather than build such programs themselves. The insignificance may be due to the fact that the effect has already been covered as stated before.

At this stage of the analysis, a reduced form model for the transaction cost study of integration is constructed. The results are consistent with the hypotheses regarding the potential holdups in the market transactions in the case of human asset specificity and uncertainty, along with the costs of managing unfamiliar or complex activities within the firm. Thus, from the above we see that the hypothesis 1 from section 3.3 is satisfied.

### **5.6 First Stage Selection Model Excluding Uncertainty**

We create the first stage selection model again, but this time without the factor “UNCERTAINTY” in order to note the effect of its exclusion on the model as a whole.

The variable “complexity” is generally used as a proxy for measuring uncertainty in transaction cost analysis. This variable has been included in the first stage model as detailed in the previous sections. Due to the specialized nature of disease management programs, another question was included in the survey to capture all effects of uncertainty on the final form chosen by the health plan. Based on the frequencies and the correlations of this variable with the other independent factors and the dependent variable, it is seen that the effect of this factor is similar to that of “complexity”, but is insignificant in the final model. Thus, it is necessary to note the model performance and accuracy with this factor removed.

### 5.6.1 Training Stage

In this stage, we use the same 80 observations of the earlier training set as shown in section 5.4.1 step 1, but without the independent variable uncertainty, in order to study the effect of its exclusion on the whole model and its accuracy. The SAS commands and data sets used are the same as detailed in section 5.4.1 step 1, with the independent factor “uncertainty” excluded. The results for the model information are as shown below in table 5.16.

#### 5.6.1 Step 1. Creating the First Stage Selection Model Excluding Uncertainty

**Table 5.16 Seven Factor Model Response, Fit Statistics and Null Hypothesis for Training Stage**

FIRST STAGE SELECTION MODEL WITH NO UNCERTAINTY

The LOGISTIC Procedure

Model Information

Data Set	WORK.TRAINING	
Response Variable	DEP	DEP
Number of Response Levels	2	
Model	binary probit	
Optimization Technique	Fisher's scoring	

**Table 5.16 (Continued)**

Number of Observations Read 80  
Number of Observations Used 80

Response Profile

Ordered Value	DEP	Total Frequency
1	1	33
2	0	47

Probability modeled is DEP=1.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	110.441	96.711
SC	112.823	115.767
-2 Log L	108.441	80.711

Testing Global Null Hypothesis: BETA=0

**Table 5.16 (Continued)**

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.73	7	0.0002
Score	23.5741	7	0.0014
Wald	18.2744	7	0.0108

**5.6.1 Step 2. Evaluate Results of the Training Stage for the Seven Factor Model**

Once the model is run, we obtain the new coefficients for the independent variables involved, which are presented in table 5.17.

**Table 5.17 Analysis of Parameter Coefficients for the Seven Factor Model**

FIRST STAGE SELECTION MODEL WITH NO UNCERTAINTY

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Standard Estimate	Error	Wald Chi – Square	Pr > ChiSq
Intercept	1	1.3695	1.4507	0.8912	0.3452
TEMPORAL	1	0.0559	0.2165	0.0667	0.7962
PHYSICAL	1	0.6194	0.2563	5.8406	0.0157
HUMAN	1	-0.5072	0.2469	4.2191	0.04
CAPITAL	1	0.5062	0.1723	8.6348	0.0033
COMPLEXITY	1	-0.6265	0.2302	7.4046	0.0065
SIMILARITY	1	-0.2675	0.1354	3.9058	0.0481
FREQUENCY	1	-0.1452	0.2136	0.462	0.4967

**Table 5.17 (Continued)**

Association of Predicted Probabilities and Observed Responses

Percent Concordant	80.6	Somers' D	0.622
Percent Discordant	18.4	Gamma	0.628
Percent Tied	1	Tau-a	0.305
Pairs	1551	C	0.811

**5.6.1 Step 3. Use the Classification Table to Determine Optimal Cut-Off Point**

As in section 5.4.1 (step 3), the classification table is constructed once again in order to obtain the best possible cut-off point for classification of the data points in the training and validation sets. Table 5.18 details the results of this step. It is seen that 55% once again serves as the best cut-off point for the classification purposes of the model. The observations from both sets that have a predicted value of less than 0.55 are classified as outsourced and the one's that have a predicted value of 0.55 or higher are classified as having an integrated organizational form.

**Table 5.18 Classification Table for the Seven Factor Training Stage**

Prob Level	Correct		Incorrect		Percentages				
	Event	Non – Event	Event	Non - Event	Correct	Sensitivity	Specificity	False POS	False NEG
0.05	33	3	44	0	45	100	6.4	57.1	0
0.1	32	11	36	1	53.8	97	23.4	52.9	8.3
0.15	32	11	36	1	53.8	97	23.4	52.9	8.3
0.2	25	17	30	8	52.5	75.8	36.2	54.5	32
0.25	24	20	27	9	55	72.7	42.6	52.9	31
0.3	23	21	26	10	55	69.7	44.7	53.1	32.3
0.35	23	25	22	10	60	69.7	53.2	48.9	28.6
0.4	23	30	17	10	66.3	69.7	63.8	42.5	25
0.45	23	37	10	10	75	69.7	78.7	30.3	21.3
0.5	20	39	8	13	73.8	60.6	83	28.6	25
0.55	17	44	3	16	76.3	51.5	93.6	15	26.7
0.6	15	45	2	18	75	45.5	95.7	11.8	28.6
0.65	15	45	2	18	75	45.5	95.7	11.8	28.6
0.7	14	45	2	19	73.8	42.4	95.7	12.5	29.7
0.75	9	45	2	24	67.5	27.3	95.7	18.2	34.8
0.8	7	45	2	26	65	21.2	95.7	22.2	36.6
0.85	7	47	0	26	67.5	21.2	100	0	35.6
0.9	7	47	0	26	67.5	21.2	100	0	35.6
0.95	4	47	0	29	63.8	12.1	100	0	38.2
1	0	47	0	33	58.8	0.0	100	.	41.3

**5.6.2 Testing Stage**

The new model is to be tested again for prediction accuracy, which done as below by running the model again using a combination of both the training and validation sets as input. The SAS commands and data sets are the same as detailed in section 5.5.1 step 1, excluding the factor “uncertainty”.

**5.6.2 Step 1. Embed the Validation Set Into the Training Set**

In the first step, the two sets (training and validation) are combined, and the model is run again as below. The results are as reported in table 5.19.

**Table 5.19 Seven Factor Model Response, Fit Statistics and Null Hypothesis for Testing Stage**

FIRST STAGE SELECTION MODEL FOR COMBINED DATA SET WITH NO  
UNCERTAINTY

The LOGISTIC Procedure  
Model Information

Data Set                    WORK.COMBINED\_NO\_UNCERT  
 Response Variable            DEP                    DEP  
 Number of Response Levels    2  
 Model                        binary probit  
 Optimization Technique       Fisher's scoring

Number of Observations Read    93  
 Number of Observations Used    80

Response Profile

Ordered Value	DEP	Total Frequency
1	1	33
2	0	47

Probability modeled is DEP=1.

NOTE: 13 observations were deleted due to missing values for the response or explanatory variables.

**Table 5.19 (Continued)**

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	110.441	96.711
SC	112.823	115.767
-2 Log L	108.441	80.711

FIRST STAGE SELECTION MODEL FOR COMBINED DATA SET WITH NO  
UNCERTAINTY

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	27.73	7	0.0002
Score	23.5741	7	0.0014
Wald	18.2744	7	0.0108

**5.6.2 Step 2. Results of the First Stage Selection Model With Combined Data Set for Seven TCE Factors**

The model run provides us with the parameter coefficients again as detailed below in table 5.20. The output parameters are similar to those explained in section 5.4.1 step 2.

**Table 5.20 Analysis of Parameter Coefficients for the Testing Stage of the Seven Factor Model**

Parameter	DF	Standard Estimate	Error	Wald Chi – Square	Pr > ChiSq
Intercept	1	1.3695	1.4507	0.8912	0.3452
TEMPORAL	1	0.0559	0.2165	0.0667	0.7962
PHYSICAL	1	0.6194	0.2563	5.8406	0.0157
HUMAN	1	-0.5072	0.2469	4.2191	0.04
CAPITAL	1	0.5062	0.1723	8.6348	0.0033
COMPLEXITY	1	-0.6265	0.2302	7.4046	0.0065
SIMILARITY	1	-0.2675	0.1354	3.9058	0.0481
FREQUENCY	1	-0.1452	0.2136	0.462	0.4967

**Table 5.20 (Continued)**

Association of Predicted Probabilities and Observed Responses

Percent Concordant	80.6	Somers' D	0.622
Percent Discordant	18.4	Gamma	0.628
Percent Tied	1	Tau-a	0.305
Pairs	1551	C	0.811

**5.6.3 Classification of the Training Set**

From section 5.6.1 step 3, we have selected 55% as the cut-off point to differentiate between integrated and contracted programs (the same as in the case of the full model), and we check the accuracy of the new model on the training set first. This is done in order to gauge the accuracy of the new model, which may change significantly due to the exclusion of the factor “uncertainty”. The diagonal elements again represent the correct classifications for both the training and the validation phase. The results for the training set are detailed in table 5.21.

**Table 5.21 Prediction Accuracy for the Training Set**

The FREQ Procedure

Table of PRED\_FORM by DEP

Frequency	0	1	Total
0	45	13	58
1	2	20	22
Total	47	33	80

Frequency Missing = 13

#### 5.6.4 Classification of the Validation Set

We also need to test the new model on the validation set in order to get the true accuracy of the model. The results are as shown below in table 5.22.

**Table 5.22 Prediction Accuracy for the Validation Set**

The FREQ Procedure  
Table of PRED\_FORM by ACTUAL\_FORM

Frequency	0	1	Total
0	5	2	7
1	1	5	6
Total	6	7	13

Frequency Missing = 80

From the above, it is seen that the prediction accuracy of the seven factor model is the same as that of the full model for both the training and validation sets. It can thus be concluded that exclusion of the factor “uncertainty” does not affect the model accuracy.

#### 5.7 Comparison of the Model With and Without the TCE Factor Uncertainty

Presented below in table 5.23 is a comparison of the parameter coefficients for both the full and seven factor models.

**Table 5.23 Comparison of the Full and Seven Factor Models**

Parameter	DF	Estimate	
		w/ Uncertainty	w/o Uncertainty
Intercept	1	1.6295	1.3695
		(0.3276)	(0.3452)
TEMPORAL	1	0.0476	0.0559
		(0.827)	(0.7962)
PHYSICAL	1	0.6327	0.6194
		(0.0144)	(0.0157)
HUMAN	1	-0.5188	-0.5072
		(0.0377)	(0.04)
CAPITAL	1	0.5214	0.5062
		(0.004)	(0.0033)
COMPLEXITY	1	-0.6031	-0.6265
		(0.012)	(0.0065)
SIMILARITY	1	-0.2788	-0.2675
		(0.0491)	(0.0481)
FREQUENCY	1	-0.1736	-0.1452
		(0.4469)	(0.4967)
UNCERTAINTY	1	-0.067	
		(0.7471)	

Pr>Chisq statistics in parenthesis

### 5.8 Inference From the Comparison of Full and Seven Factor Model

It is seen that the coefficient and significance of “TEMPORAL” increases slightly, while the factor coefficient of “COMPLEXITY” decreases. This effect is consistent with the observed correlations between the variables for temporal specificity and uncertainty as “TEMPORAL” is negatively correlated with the factor “UNCERTAINTY” due to which its value increases when the second factor for uncertainty is removed. The coefficients for factors “SIMILARITY”, “FREQUENCY” and “HUMAN” also show the same effect. On the other hand, “COMPLEXITY” is positively correlated with “UNCERTAINTY”, and thus the value of this factor decreases upon the removal of “UNCERTAINTY”. The same can be said for the values of coefficients for the factors “PHYSICAL” and “CAPITAL”. Thus, we see that the coefficients of the factors change slightly, however the prediction accuracy of the model remains unchanged.

## **5.9 Calculation of Actual In-Sourced and Outsourced Costs**

Based on the literature review and industrial inquiry, the average hourly wage for health plan executives and decision makers responsible for initiating, developing and managing DM programs was found to be \$60. This value was multiplied to the time estimates that were obtained from our survey in order to obtain an estimate of the actual transaction costs for each DM program, the results of which have been presented below.

### **5.9.1 Breakup of Actual In-Sourced Costs**

Using the time estimates from the survey, and the average hourly wage, we obtain the estimates of the transaction costs for both the in-sourced and outsourced cases of the responses. The means and statistics for the in - sourced case are reported in table 5.24.

For the in-sourced costs, the questions 3a (time taken for administrative/facility planning tasks), and 4 (search and information time spent) are multiplied with the number of hours in each working day (taken here as 8 hours/day) and then multiplied with the average hourly management wage, which is \$60. To this is added the supervisory cost, which is given in terms of hours per week. To annualize it, the number of weeks in a year is multiplied along with the average hourly management wage in order to obtain this cost. The sum of these three elements gives us an estimate if the in-sourced costs for a particular integrated disease management program.

For the estimation of the outsourced costs, the cost estimates from questions 4 and 5 are calculated and added as above; in addition, the value obtained from question 3a (legal/negotiations costs) is directly added to the above as a direct dollar amount is asked for this question, and no conversion is necessary. The addition of these three values provides the outsourced cost estimate for each observation in the outsourced subset.

The first row of table 5.24 details the administrative/facility planning costs for the integrated cases of disease management programs, while the second and third rows show the search and information costs and supervisory costs respectively. The last row reports the means and statistics for the total in-sourced costs, which is the sum of the first three rows.

**Table 5.24 Means for the Actual In-Sourced Costs**

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
INSOURCED_ADMIN_COST	40	55812	68014.12	0	288000
SEARCH_INFO_COST	40	27408	47821.8	0	288000
SUPERVISORY_COST	40	52488	68036.08	0	288000
TOTAL_INSOURCED_COST	40	135708	150068.67	0	691200

### 5.9.2 Breakup of Actual Outsourced Costs

As in the earlier case, the means and statistics for the outsourced case are presented in table 5.25 below. The first row shows the legal cost, which is in dollars. The second and third rows report the dollar values of the calculated search and information costs and the supervisory costs, while the fourth row reports the means and statistics for the total outsourced costs, which is the sum of the first three terms explained above.

**Table 5.25 Means for Actual Outsourced Costs**

The MEANS Procedure

Variable	N	Mean	Std Dev
LEGAL_COST_OUTSOURCED_DOLLARS	53	14132.08	24255.99
SEARCH_INFO_COST	53	17750.94	21043.81
SUPERVISORY_COST	53	20513.21	21706.79
TOTAL_OUTSOURCED_COST	53	52396.23	48103.72
Variable	Minimum	Maximum	
LEGAL_COST_OUTSOURCED_DOLLARS	0	100000	
SEARCH_INFO_COST	0	86400	
SUPERVISORY_COST	0	108000	
TOTAL_OUTSOURCED_COST	0	197920	

We see that outsourced costs are approximately 1/3 of in-sourced costs due to the difficulties noted by Masten et al., who state that contracting costs are incurred by each party included in the transaction, hence, cost data needs to be collected from two or more sources. In addition, the contractual changes and failures that occur in this case occur probabilistically over time, which requires that data be collected on the intangible expectations of the decision makers. Thus, the collected outsourcing costs have been disregarded and the focus is placed on the in-sourced costs for building of the predictive models in the next stage.

### **5.9.3 Calculation of the Inverse Mills Ratio and the Help and Control Factor Delta**

As the outsourced are disregarded, the dependent variable information is missing for part of our data set. The standard selection bias problem is thus encountered when constructing the cost equation as detailed in chapter 4. In order to correct the selection bias, an additional independent variable is needed to be added along with the transaction cost factors. This variable is the Heckman correction factor lambda, which is the inverse mills ratio.

To compute the Heckman correction factor Lambda with a PROBIT selection model, the following SAS commands are used:

```

proc logistic data=COMBINED descending;
model DEP = temporal physical human capital complexity
similarity frequency uncertainty/LINK=PROBIT ;
output out=prob2 XBETA= g2 predicted=phat2;
TITLE 'FIRST STAGE SELECTION MODEL FOR COMBINED DATA SET';
run;

```

Where, “combined” is the data set containing all 93 data points, the command “XBETA” gives us the probit scores generated for each of the observations. The command “predicted” provides us with the probabilities for each of the observations recorded.

In the output of this analysis, we find the estimates of the parameters. On the basis of these parameters, for each observation the predicted probit score is also obtained, which is stored in the variable “g2”. These probit scores obtained in the variable “g2” are used to compute the Heckman control factor LAMBDA, using the SAS command as follows:

$$\text{LAMBDA1} = ((1/\sqrt{2*3.141592654})*(\exp(-G2*G2*0.5)))/\text{CDF}(\text{'NORMAL'},G2);$$

Or

$$\text{LAMBDA2} = (\text{PDFG2}/\text{CDFG2});$$

Or

$$\text{lambda3} = (1/\sqrt{2*3.141592654})*\exp(-1*g2**2/2)/\text{probnorm}(g2).$$

For applying the two-step procedure it is important that all rows with missing values on variables which are used in the substantial analyses are removed from the active file. This means that all the outsourced cases and the cases where the dependent variable or any of the independent variables are missing are removed, and the following analysis is done on the remaining in-sourced subset only. The next step is to compute the value of the control factor:

$$\text{DELTA1} = -\text{LAMBDA1}*G2-\text{LAMBDA1}*\text{LAMBDA1};$$

$$\text{DELTA2} = -\text{LAMBDA2}*G2-\text{LAMBDA2}*\text{LAMBDA2};$$

$$\text{DELTA3} = -\text{LAMBDA3}*G2-\text{LAMBDA3}*\text{LAMBDA3};$$

Three values of the control factor are calculated in order to check the values of all three inverse mills ratios obtained by the different methods. The values of DELTA1, DELTA2 and DELTA3 should be between -1 and 0. The values of both the inverse mills ratio and the control factor are checked as follows:

```

PROC MEANS DATA = LAMBDA ;
VAR LAMBDA1 LAMBDA2 lambda3 DELTA1 DELTA2 DELTA3 H1 H2 H3;
TITLE 'RESULTS FOR THE INVERSE MILLS RATIO AND CONTROL FACTOR DELTA';
RUN;

```

The table 5.26 details the values of the inverse mills ratio and the control factor delta. “Data = Lambda” denotes the data set lambda, from which all outsourced and missing data points have been excluded.

**Table 5.26 Results and Statistics for the Inverse Mills Ratio and Control Factor Delta**

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
LAMBDA1	39	0.6465018	0.4819293	0.0085964	1.7400238
LAMBDA2	39	0.6465018	0.4819293	0.0085964	1.7400238
lambda3	39	0.6465018	0.4819293	0.0085964	1.7400238
DELTA1	39	-0.4992273	0.2347647	-0.8291143	-0.0238978
DELTA2	39	-0.4992273	0.2347647	-0.8291143	-0.0238978
DELTA3	39	-0.4992273	0.2347647	-0.8291143	-0.0238978
h1	39	0.4992273	0.2347647	0.0238978	0.8291143
h2	39	0.4992273	0.2347647	0.0238978	0.8291143
h3	39	0.4992273	0.2347647	0.0238978	0.8291143

All three formulas give us the same values for the variable “lambda”. The variables “h1”, “h2”, and “h3” are the inverses of the three control factors calculated. The value of the control factor delta (DELTA1 in our case) should be between -1 and 0 which is satisfied for all three cases as seen above. Hence, one case of the calculated inverse mills ratio and the control factor delta can be used in the planned analysis. The inverse mills ratio is calculated and added to the analysis as an additional independent variable as detailed in chapter 4.

## **5.10 Frequencies for the In-Sourced Subset**

Before constructing the cost model, the frequencies of the integrated portion of the data set is presented.

### **5.10.1 Frequencies for the Organizations in the In-Sourced Subset**

Table 5.27 reports the frequencies for the health plans that have in-house DM programs, whereas table 5.28 in the next section provides the number of DM program present in this subset of the full data set.

**Table 5.27 Frequencies for Responding Organizations of the In-Sourced Subset**

The FREQ Procedure

Organization	Frequency	Cumulative Frequency
Ault International Medical Management, LLC	1	1
CareGuide, Inc.	5	6
Contra Costa Health Plan	1	7
Florida Health Care Plans	4	11
Health Alliance Plan	4	15
HealthPartners	2	17
Healthy Futures, Inc	1	18
IMS Managed Care, Inc.	5	23
Memphis Managed Care Corp	1	24
Miller & Huffman Outcome Architects, LLC	2	26
Mountain States Home Care	1	27
Partners HealthCare	1	28
QualChoice	5	33
Quality First Healthcare, Inc.	1	34
Solucia Inc	5	39
WellPoint, Inc.	1	40

**5.10.2 Frequencies for the Diseases in the In-Sourced Subset**

The frequencies for the diseases managed by the programs implemented in-house by the health plans in the in-sourced subset are as shown below in table 5.28.

**Table 5.28 Frequencies for the DM Programs of the In-Sourced Subset**

Disease	Frequency	Cumulative Frequency
Asthma	7	7
Chronic Obstructive Pulmonary Disease (COPD)	4	11
Congestive Heart Failure (CHF)	8	19
Coronary Artery Disease(CAD)	7	26
Diabetes	10	36
Other: Hypertension	2	38
Other: all chronic health conditions	1	39
Pressure ulcers	1	40

**5.11 Organization Cost Model for In-Sourced Costs**

In this analysis, the factor “UNCERTAINTY” is removed, as this factor is a proxy for the transaction cost factor uncertainty, which is already accounted for with the variable “COMPLEXITY”. Also, as detailed by Heckman, one independent variable from the selection equation must be removed during the substantial analysis. Inclusion of all eight independent variables from the first stage into the organization cost equation causes the estimated correlation coefficients between the errors in the selection and cost equations to exceed the logical upper bound. The SAS commands used are given below:

```

PROC REG DATA=LAMBDA;
MODEL TOTAL_INSOURCED_COST = temporal physical human capital complexity
similarity frequency LAMBDA1;
output out=INSOURCED_PRED predicted=PRED_IN_COST;
TITLE 'ORGANIZATION COST MODEL FOR IN - SOURCED COSTS';
RUN;

```

In this analysis, the in-sourced cost (total\_insourced\_cost) is the dependent variable, whereas the transaction cost factors (excluding “uncertainty”) and the inverse mills ratio are the independent variables. The data set “lambda” is that part of the combined data set that has no missing independent or dependent variables and contains only the integrated portion of the total sample. The command “predicted” produces the predicted values of the in-source costs and stores them in the variable called “pred\_in\_cost” in the data set “insourced\_pred” (defined by the “output out” command). The results are as detailed in the sections below.

### **5.11.1 Running the Model**

In this step, the model is run with the calculated in-house cost as the dependent variable, and the survey responses to the TCE questions as the independent variables. The tests and statistics displayed in this part of the results are explained below:

- 1) Source - The source of variance, Model, Residual, and Total. The Total variance is divided into the variance which can be explained by the independent variables (Model) and the variance which is not explained by the independent variables (Residual or Error).
- 2) DF - The degrees of freedom associated with the sources of variance. The total variance has  $N-1$  degrees of freedom. In this case,  $N=39$ , so the DF for total is 38. The model degrees of freedom correspond to the number of predictors minus 1 ( $K-1$ ). The intercept is automatically included in the model. Including the intercept, there are 9 predictors, so the model has  $9-1=8$  degrees of freedom. The Residual degree of freedom is the DF total minus the DF model,  $38 - 8$  is 30.
- 3) Sum of Squares - The Sum of Squares associated with the three sources of variance, Total, Model and Residual.
- 4) Mean Square - The Sum of Squares divided by their respective DF.
- 5) F Value and Pr > F - The F-value is the Mean Square Model divided by the Mean Square Residual. The p-value associated with this F value is displayed. The p-value is compared to the alpha level (typically 0.05) and, if smaller, it can be concluded that the independent variables reliably predict the dependent variable. If the p-value is greater than 0.05, it can be said that the group of independent variables does not show a statistically significant relationship with the dependent variable, or that the group of independent variables does not reliably predict the dependent variable.

- 6) Root MSE - Root MSE is the standard deviation of the error term, and is the square root of the Mean Square Residual (or Error).
- 7) Dependent Mean - The mean of the dependent variable.
- 8) Coeff Var - The coefficient of variation, which is a unit-less measure of variation in the data. It is the root MSE divided by the mean of the dependent variable.
- 9) R-Square - R-Square is the proportion of variance in the dependent variable (in-sourced cost) which can be predicted from the independent variables (the transaction cost factors and inverse mills ratio). This value indicates that 68.65% of the variance in science scores can be predicted from the independent variables.
- 10) Adj R-Sq - As predictors are added to the model, each predictor explains some of the variance in the dependent variable simply due to chance. The adjusted R-square attempts to yield a more honest value to estimate the R-squared for the population. The value of R-square was 0.6865, while the value of Adjusted R-square was 0.6029. Adjusted R-squared is computed using the formula  $1 - ((1 - Rsq)(N - 1) / (N - k - 1))$ .

Table 5.29 reports the results from the analysis of variance and the R – square value for the model constructed.

**Table 5.29 Analysis of Variance for the In-Sourced Cost Model**

The REG Procedure

Model: MODEL1

Dependent Variable: TOTAL\_INSOURCED\_COST

Number of Observations Read 40  
 Number of Observations Used 39  
 Number of Observations with Missing Values 1

Source	DF	Sum of Squares	Mean Square Value	F	Pr > F
Model	8	5.900092E+11	73751145750	8.21	<.0001
Error	30	2.694055E+11	8980184083		
Corrected Total	38	8.594147E+11			

Root MSE	94764	R-Square	0.6865
Dependent Mean	139188	Adj R-Sq	0.6029
Coeff Var	68.08349		

**5.11.2 Results of the In-Sourced Cost Model With Correction for Selection Bias**

Table 5.30 below reports the coefficients of the independent factors for the linear specification of the internal cost equation. The explanation for the terms and statistics found in this result are given below:

- 1) Variable - This column shows the predictor variables (the independent transaction cost factor). The first variable represents the intercept.
- 2) Label - This column gives the label for the variable.
- 3) DF - This column give the degrees of freedom associated with each independent variable.
- 4) Parameter Estimates - The values for the regression equation for predicting the dependent variable from the independent variable. The regression equation is presented in many different ways, for example:

$$Y_p = a_0 + a_1*x_1 + a_2*x_2 + a_3*x_3 + a_4*x_4 + \dots + a_N*x_N$$

The column of estimates provides the values for  $a_0$ ,  $a_1$ ,  $a_2$ , and so on for this equation. In this case, the regression equation is:

$$\begin{aligned} \text{TOTAL\_INSOURCED\_COST} = & 277864 + (-123102)*\text{TEMPORAL} + (- \\ & 71011)*\text{PHYSICAL} + 142950*\text{HUMAN} + (-98136)*\text{CAPITAL} + \\ & 18113*\text{COMPLEXITY} + 53406*\text{SIMILARITY} + 160158*\text{FREQUENCY} + (- \\ & 335922)*\text{LAMBDA} \end{aligned}$$

- 5) Standard Error - The standard errors associated with the coefficients. The standard errors in this case are still biased, which shall be removed using the methodology detailed in section 5.11.3.
- 6) t Value and Pr > |t| - These columns provide the t-value and 2 tailed p-value used in testing the null hypothesis that the coefficient/parameter is 0. Coefficients having p-values less than alpha are statistically significant.

**Table 5.30 Parameter Estimates for the Independent Variables in the In-Sourced Cost Model**

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	Intercept	1	277864	129379	2.15	0.0399
TEMPORAL	TEMPORAL	1	-123102	22140	-5.56	<.0001
PHYSICAL	PHYSICAL	1	-71011	67213	-1.06	0.2992
HUMAN	HUMAN	1	142950	48164	2.97	0.0058
CAPITAL	CAPITAL	1	-98136	34980	-2.81	0.0087
COMPLEXITY	COMPLEXITY	1	18113	45284	0.4	0.692
SIMILARITY	SIMILARITY	1	53406	22626	2.36	0.025
FREQUENCY	FREQUENCY	1	160158	27855	5.75	<.0001
LAMBDA1		1	-335922	139581	-2.41	0.0225

### 5.11.3 Correcting the Standard Error Terms

The above analysis produces unbiased parameter estimates for the independent variables. However, the standard estimates of these parameters are biased because of heteroskedasticity. The variance of the error term is not the same for each respondent. To correct the standard errors and get the unbiased estimates, the following additional steps have to be taken.

First, a command was added to the substantial regression analysis to save the residuals of the regression model in a new variable (which is called RES), as given below:

```
PROC REG DATA=LAMBDA;  
MODEL TOTAL_INSOURCED_COST = temporal physical human capital complexity  
similarity frequency LAMBDA1;  
output out=INSOURCED_PRED predicted=PRED_IN_COST residual= RES;  
RUN;
```

This variable must be squared:

```
RES2 = RES*RES.
```

Two other help variables must also be computed. The first one is the regression coefficient of the variable LAMBDA (the inverse mills ratio) in the OLS analysis, which is called LAMB. The second one is the number of cases used in the OLS regression, called N. The results for both are given below.

```
LAMB=-335922.
```

```
N=39.
```

The variable RES2 and also DELTA, which was computed in the first part of the analysis, have to be summed over all cases. The values of the sum of these two variables are as given below.

```
DELTAS1 = -19.4699.
```

```
RESS = 2.694E11.
```

Where, RESS and DELTAS1 are the sums of the residuals and the control factor DELTA, respectively.

Now the corrected value of the variance (VARC) and the standard error (SEC) of the error term of the substantial equation can be estimated:

```
VARC = RESS/N-LAMB*LAMB*DELTAS/N.
```

$$SEC = \sqrt{VARC}.$$

Computation of RHO, the correlation between the error terms of the selection and substantial equations:

$$RHO = \sqrt{LAMB * LAMB / VARC}.$$

If (lamb < 0) RHO = 0 - RHO.

Now the values of VARC, SEC and RHO can be computed, the values of which are noted in table 5.31:

**Table 5.31 Values of Corrected Variance, Std. Error and Error Correlation**

The MEANS Procedure

Variable	N	Value
VARC	40	63242395246
SEC	40	251480.41
RHO	40	-1.335778

Computation of the standard errors of the separate observations (RHOI) and transformation of the standard errors into weights (WGT):

$$RHOI = \sqrt{VARC + LAMB * LAMB * DELTA}.$$

$$WGT = 1/RHOI.$$

Now the corrected standard errors can be computed by running the substantial analysis again, but this time as Weighted Least Squares (WLS) regression with WGT as weight:

**PROC REG DATA=INSOURCED\_PRED;**

**MODEL** TOTAL\_INSOURCED\_COST = temporal physical human capital complexity  
similarity frequency LAMBDA1;

**weight** WGT;

```

output out=INSOURCED_PRED_NEW predicted=PRED_IN_COST_NEW residual=
RES_NEW;
title 'CORRECTING HETEROSKEDASTICITY FOR COST MODEL';
RUN;

```

The results of the new regression are as reported below in table 5.32. The term “weight” indicates that the variable “WGT” calculated above is used as a weight in this regression.

**Table 5.32 Results From Heteroskedasticity Correction**

CORRECTING HETEROSKEDASTICITY FOR COST MODEL

The REG Procedure

Model: MODEL1

Dependent Variable: TOTAL\_INSOURCED\_COST

Number of Observations Read				40	
Number of Observations Used				39	
Number of Observations with Missing Values				1	
Weight: WGT					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	4964233	620529	6.58	<.0001
Error	30	2827930	94264		
Corrected Total	38	7792163			
Root MSE					
Root MSE		307.025	R-Square	0.6371	
Dependent Mean		141982	Adj R-Sq	0.5403	
Coeff Var		0.21624			

**Table 5.32 (Continued)**  
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	Intercept	1	196189	127481	1.54	0.1343
TEMPORAL	TEMPORAL	1	-105395	24307	-4.34	0.0002
PHYSICAL	PHYSICAL	1	-70407	75141	-0.94	0.3562
HUMAN	HUMAN	1	129803	55543	2.34	0.0263
CAPITAL	CAPITAL	1	-81541	40472	-2.01	0.053
COMPLEXITY	COMPLEXITY	1	25955	54356	0.48	0.6365
SIMILARITY	SIMILARITY	1	38124	25757	1.48	0.1493
FREQUENCY	FREQUENCY	1	149350	30494	4.9	<.0001
LAMBDA1		1	-280786	163067	-1.72	0.0954

By combining the parameter estimates of the substantial analysis with the standard errors of this WLS analysis, the Heckman procedure is completed. To indicate the explained variance  $R^2$  of the analysis, the  $R^2$  of the substantial analysis should be taken. Thus, combining the parameter estimates and  $R^2$  from the initial step and the corrected standard errors, we get the final results as shown in table 5.33. The standard errors given below have been corrected for heteroskedasticity and endogeneity of the selection correction index.

**Table 5.33 Final Cost Model Results**

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	Intercept	1	277864	127481	2.15	0.0399
TEMPORAL	TEMPORAL	1	-123102	24307	-5.56	<.0001
PHYSICAL	PHYSICAL	1	-71011	75141	-1.06	0.2992
HUMAN	HUMAN	1	142950	55543	2.97	0.0058
CAPITAL	CAPITAL	1	-98136	40472	-2.81	0.0087
COMPLEXITY	COMPLEXITY	1	18113	54356	0.4	0.692
SIMILARITY	SIMILARITY	1	53406	25757	2.36	0.025
FREQUENCY	FREQUENCY	1	160158	30494	5.75	<.0001
LAMBDA1		1	-335922	163067	-2.41	0.0225

$$R^2 = 0.6865$$

#### 5.11.4 Inference From the Internal Cost Equation

The second stage results above from the internal organization cost model confirm and strengthen the predictions of the theory and the findings of the first stage selection model with regard to the effects of TCE factors on the sourcing decisions for health plans.

The effect of temporal specificity (TEMPORAL) on the integration transaction costs is negative, as meaning that health plans will tend to reduce their transaction costs if DM programs that require stricter adherence to timing and scheduling are built in-house rather than contracted. The significance of the factor indicates that the effect of this factor fosters integration through its effect on internal organization costs rather than by increasing the hazards of market exchange, as

noted by Masten et al. The effect of physical asset specificity (PHYSICAL) on the integration transaction costs is also negative, as meaning that health plans will tend to reduce their transaction costs if DM programs that require more specific tools and software are integrated rather than outsourced. However, the coefficient on this factor is not significant, indicating that the principal effect of PHYSICAL on the integration decision derives from the hazards of market exchange.

The cost coefficient of human asset specificity (HUMAN) is positive in the model, meaning that if DM programs requiring specific skills and knowledge from its employees are in-sourced, transaction costs for the health plan tend to rise. The second-stage estimates indicate that the correlation between the human asset specificity and the likelihood of integration is a consequence of the rise in internal organization costs, rather than decrease in costs of market exchange as the theory predicts.

The coefficient of dedicated asset specificity (CAPITAL) is negative, meaning that if DM programs requiring greater investments unique to the program are in-sourced, transaction costs for the health plan tend to decrease. The significance of the factor again indicates that the effect of this factor fosters integration through its effect on internal organization costs rather than by increasing the hazards of market exchange, as with the factors for temporal and human asset specificity.

The effect of uncertainty (COMPLEXITY) is positive, meaning that DM programs for which effectiveness and performance measurement are more difficult should be outsourced by health plans in order to reduce their incurred transaction costs. The coefficient on this factor is not significant, indicating that the principal effect of uncertainty on the integration decision derives from the hazards of market exchange. The effect of similarity (SIMILARITY) on the in-sourced transaction cost is also positive, meaning that if DM programs similar to the ones the health plan may be involved in are integrated, transaction costs for the health plan tend to rise. Unlike complexity, this factor is also significant, indicating that the effect of this factor fosters integration through its effect on internal organization costs rather than by increasing the hazards of market exchange.

Finally, the effect of frequency (FREQUENCY) in the in-sourced transaction cost model is positive, meaning that if DM programs health plans risk increasing their incurred transaction costs if they integrate DM programs requiring a high frequency of contact with the individuals enrolled in the program. The significance of the factor indicates that the effect of this factor fosters integration through its effect on internal organization costs, and not by increasing the hazards of market exchange.

The second stage findings confirm and strengthen the findings of the first stage estimation with regards to the transaction cost factors. In addition, we can deduce that the factors similarity, temporal, human, capital and frequency have their primary effect on the internal organization costs rather than on market costs as the theory suggests, whereas the factors complexity and physical act principally on the costs of market exchange. Thus, from the above, it is seen that hypothesis 2 stated in section 3.3 is satisfied.

## **5.12 Comparison of First and Second Stage Results**

Table 5.34 presents a comparison between the coefficients obtained for the independent variables from the selection and the substantial equation. A side by side comparison of the parameter coefficients obtained from the first and second stage parameter coefficients establishes the fact that the effect of the transaction cost factors is captured both in terms of effect on organizational form and in terms of costs in the case of health plans.

**Table 5.34 Comparison of Selection and Substantial Model Coefficients**

Parameter	DF	Estimate	
		First stage	Second stage cost estimate
Intercept	1	1.6295 (0.3276)	277864 (0.0399)
TEMPORAL	1	0.0476 (0.8270)	-123102 (<.0001)
PHYSICAL	1	0.6327 (0.0144)	-71011 (0.2992)
HUMAN	1	-0.5188 (0.0377)	142950 (0.0058)
CAPITAL	1	0.5214 (0.0040)	-98136 (0.0087)
COMPLEXITY	1	-0.6031 (0.0120)	18113 (0.6920)
SIMILARITY	1	-0.2788 (0.0491)	53406 (0.0250)
FREQUENCY	1	-0.1736 (0.4469)	160158 (<.0001)
UNCERTAINTY	1	-0.067 (0.7471)	
LAMBDA (inverse mills ratio)	1		-335922 (0.0225)

Pr>Chisq statistics in parenthesis

### 5.13 Comparison of Actual and Predicted In-Sourced Costs

In order to determine the accuracy and effectiveness of the model, we need to check the means and statistics of the predicted costs with the actual recorded values. Table 5.35 presents the means for both the actual and predicted costs, while table 5.36 reports the predicted value and the error for a sub-sample of the in-sourced set.

#### 5.13.1 Means for the Actual and Predicted In-Sourced Costs

From table 5.35, the mean, standard deviation and the minimum and maximum values for the actual and predicted in-sourced costs can be inferred. The variable “total\_inourced\_cost” is the actual transaction cost of setting up and maintaining a disease management program in-house, whereas the variable “pred\_in\_cost” is the predicted cost produced by the second stage regression model as detailed in the earlier section.

**Table 5.35 Means and Statistics for Actual and Predicted In-Sourced Costs**

Variable	N	Mean	Std Dev
TOTAL_INSOURCED_COST	40	135708	150068.67
PRED_IN_COST	39	139187.69	124605.61
Variable	Minimum	Maximum	
TOTAL_INSOURCED_COST	0	691200	
PRED_IN_COST	-52659.15	585426.79	

### 5.13.2 Comparison of Actual and Predicted Costs for Sub-Sample of Data

Presented below in table 5.36 is a sub-sample of the integrated subset from which the prediction accuracy can be determined. Also calculated is the prediction error, which is also reported for the chosen subset. The error for each case is calculated as the (actual cost- predicted cost)/(actual cost)\*100.

**Table 5.36 Comparison Between Actual and Predicted Costs for Sub-Sample of Data**

Organization	Disease	ACTUAL COST (\$)	PREDICTED COST (\$)	ERROR (%)
Florida Health Care Plans	Other: Hypertension	178,560	192,234.7674	-7.658359877
Partners HealthCare	Congestive Heart Failure (CHF)	691,200	585,426.7885	15.30283731
Health Alliance Plan	Diabetes	376,800	294,539.1615	21.83143271
Ault International Medical Management, LLC	Other: all chronic health conditions	288,000	209,266.0364	27.33818179
Memphis Managed Care Corp	Diabetes	333,600	216,655.5245	35.05529842

### 5.13.3 Rolling Up the Costs for Each Organization in the In-Sourced Subset

Since most organizations in the integrated subset have multiple DM programs, we can combine the costs for each program to get a total value for each organization. Table 5.37 presents the means of the total costs (actual and predicted) for each organization. The variable “sum\_actual\_in\_costs” represents the sum of the integrated costs for each disease management program offered by a particular health plan. The second variable (sum\_pred\_in\_cost) is the sum

of the predicted in-house costs for each disease management program that were obtained from the second stage cost model for each given organization.

**Table 5.37 Rolling Up the Costs for Each Organization in the Integrated Subset**

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
SUM_ACTUAL_IN_COST	16	339270	386307.2	0	1507200
SUM_PRED_IN_COST	15	361888	322802.46	-263295.75	981298.71

#### 5.14 Creating the Log Specification Model for the In-Sourced Costs

From above we see that the predicted costs are negative for a few data points. In order to constrain them in the positive direction and also to provide a better fit to the data, the log specification of the model to predict the in-sourced costs of a single DM program is taken. The results are as below:

##### 5.14.1 Running the Model

For this case, the log value of the recorded transaction costs is calculated and used as the dependent variable, whereas the seven independent transaction cost factors and the inverse mills ratio (lambda) are kept unchanged and used as the independent variables for the following model. The SAS commands for this stage are given below.

```
DATA LAMBDA_LOG;
SET LAMBDA;
IF TOTAL_INSOURCED_COST ^= 0 THEN TOTAL_INSOURCED_COST_LOG =
LOG(TOTAL_INSOURCED_COST);
RUN;
```

The data set “lambda” is the data set that was originally used to construct the second stage cost model for the in-source subset. The log value of the actual costs are now used as the dependent variable, whereas the transaction cost factors and the inverse mills ratio are used as the independent factors as before. The “predicted” command produces the predicted value of the costs and stores it in the variable “pred\_in\_cost\_log” in the new data set called “insourced\_pred\_log”.

```

PROC REG DATA=LAMBDA_LOG;
MODEL TOTAL_INSOURCED_COST_LOG = temporal physical human capital complexity
similarity frequency LAMBDA1;
output out=INSOURCED_PRED_LOG predicted=PRED_IN_COST_LOG;
RUN;

```

The results have been reported in table 5.38, and contain the same terms and statistics as explained in section 5.1.1.1. There is one observation that has missing data, which is removed, as in the earlier case and can be seen in the results below.

**Table 5.38 Analysis of Variance for Log Specification of In-Sourced Cost Model**

The REG Procedure	
Model: MODEL1	
Dependent Variable: TOTAL_INSOURCED_COST_LOG	
Number of Observations Read	40
Number of Observations Used	39
Number of Observations with Missing Values	1

**Table 5.38 (Continued)**

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	45.82037	5.72755	26.71	<.0001
Error	30	6.43346	0.21445		
Corrected Total	38	52.25382			

Root MSE	0.46309	R-Square	0.8769
Dependent Mean	11.26964	Adj R-Sq	0.844
Coeff Var	4.10915		

**5.14.2 Results of the Log Specification In-Sourced Cost Model**

Table 5.39 reports the coefficients for the log specification of the internal cost model. For an explanation of the terms please see section 5.11.2.

**Table 5.39 Parameter Estimates for the In-Sourced Log Specification Model**

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	Intercept	1	12.64238	0.63224	20	<.0001
TEMPORAL	TEMPORAL	1	-1.04877	0.10819	-9.69	<.0001
PHYSICAL	PHYSICAL	1	-1.4895	0.32845	-4.53	<.0001
HUMAN	HUMAN	1	1.58807	0.23536	6.75	<.0001
CAPITAL	CAPITAL	1	-0.50171	0.17094	-2.93	0.0063
COMPLEXITY	COMPLEXITY	1	0.61511	0.22129	2.78	0.0093
SIMILARITY	SIMILARITY	1	0.54914	0.11057	4.97	<.0001
FREQUENCY	FREQUENCY	1	0.99731	0.13612	7.33	<.0001
LAMBDA1		1	-3.33719	0.6821	-4.89	<.0001

### 5.14.3 Correcting the Standard Errors for the Log Spec Model

As before, this analysis produces unbiased parameter estimates for the independent variables. However, the standard estimates of these parameters are again biased because of heteroskedasticity, and the variance of the error term is not the same for each respondent. To correct the standard errors and get the unbiased estimates, we follow the same steps as outlined in section 5.11.3.

The values for the variables RESS and DELTAS1 and the corrected value of the variance (VARC), the standard error (SEC) of the error term of the substantial equation, and RHO, the correlation between the error terms of the selection and substantial equations is as calculated as before and is as shown below, and the values for the corrected variance, standard error and correlation are noted in table 5.40:

DELTAS1 = -19.4699

RESS = 6.433455

**Table 5.40 Log Specification Corrected Variance, Std. Error and Error Correlation**

The MEANS Procedure

Variable	N	Value
VARC	40	5.7247836
SEC	40	2.392652
RHO	40	-1.3947661

Computation of the standard errors of the separate observations (RHOI) and transformation of the standard errors into weights (WGT):

$$RHOI = \sqrt{VARC + LAMB * LAMB * DELTA}.$$

$$WGT = 1/RHOI.$$

Now the substantial analysis is run again, with the computed weights (WGT) as weight. The results of the new regression are as reported below in table 5.41.

**Table 5.41 Log Spec Heteroskedasticity Correction Results**

Number of Observations Read		40			
Number of Observations Used		39			
Number of Observations with Missing Values		1			
Weight: WGT					
Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	8	38.31308	4.78914	24.4	<.0001
Error	30	5.88712	0.19624		
Corrected Total	38	44.20021			
Root MSE					
Root MSE		0.44299	R-Square		0.8668
Dependent Mean		11.28769	Adj R-Sq		0.8313
Coeff Var		3.92451			

**Table 5.41 (Continued)**

## Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	Intercept	1	12.48772	0.59526	20.98	<.0001
TEMPORAL	TEMPORAL	1	-1.00912	0.11035	-9.14	<.0001
PHYSICAL	PHYSICAL	1	-1.38213	0.34778	-3.97	0.0004
HUMAN	HUMAN	1	1.53842	0.25731	5.98	<.0001
CAPITAL	CAPITAL	1	-0.47016	0.18438	-2.55	0.0161
COMPLEXITY	COMPLEXITY	1	0.55885	0.24576	2.27	0.0303
SIMILARITY	SIMILARITY	1	0.53643	0.12081	4.44	0.0001
FREQUENCY	FREQUENCY	1	0.92685	0.13861	6.69	<.0001
LAMBDA1		1	-3.13836	0.75417	-4.16	0.0002

Now combining the parameter estimates of the substantial analysis with the standard errors of this WLS analysis, the Heckman procedure is completed. Thus, combining the parameter estimates and  $R^2$  from the initial step and the standard errors corrected for heteroskedasticity and endogeneity of the selection correction index, the final results as shown below in table 5.42.

**Table 5.42 Final Log Spec Cost Model Results**

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr >  t
Intercept	Intercept	1	12.64238	0.59526	20	<.0001
TEMPORAL	TEMPORAL	1	-1.04877	0.11035	-9.69	<.0001
PHYSICAL	PHYSICAL	1	-1.4895	0.34778	-4.53	<.0001
HUMAN	HUMAN	1	1.58807	0.25731	6.75	<.0001
CAPITAL	CAPITAL	1	-0.50171	0.18438	-2.93	0.0063
COMPLEXITY	COMPLEXITY	1	0.61511	0.24576	2.78	0.0093
SIMILARITY	SIMILARITY	1	0.54914	0.12081	4.97	<.0001
FREQUENCY	FREQUENCY	1	0.99731	0.13861	7.33	<.0001
LAMBDA1		1	-3.33719	0.75417	-4.89	<.0001

$$R^2 = 0.876$$

### 5.15 Comparison of Actual and Predicted In-Sourced Costs

The predicted costs for the log specification of the cost model are obtained by taking the exponential of the predicted values obtained from the results as detailed in section 5.14.2.

From the above model, the predicted values of the in-sourced TCE costs are calculated again, along with the prediction error which is given as:  $((\text{actual cost} - \text{predicted cost}) / \text{actual cost}) * 100$ , with the variable name “error\_act”. The results are as given below. The SAS commands are as follows:

```
DATA INSOURCED_PRED_LOG;
SET INSOURCED_PRED_LOG;
PRED_IN_COST_NEW = EXP(PRED_IN_COST_LOG);
ERROR1_ACT = ((TOTAL_INSOURCED_COST -
PRED_IN_COST_NEW)/TOTAL_INSOURCED_COST)* 100;
RUN;

PROC MEANS DATA = INSOURCED_PRED_LOG;
VAR TOTAL_INSOURCED_COST PRED_IN_COST_NEW PRED_IN_COST_LOG
ERROR1_ACT ;
TITLE 'MEANS FOR WHOLE IN-SOURCED ORGANIZATION SET - LOG SPEC';
RUN;
```

### **5.15.1 Means for the Actual and Predicted In-Sourced Costs From the Log Specification Model**

The means, standard deviation and minimum and maximum values of the actual and predicted costs from the log specification of the model produced by the SAS commands stated above are presented in table 5.43 along with the statistics for the prediction error.

**Table 5.43 Means and Statistics for Actual and Predicted In-Sourced Costs From the Log Specification Model**

The MEANS Procedure

Variable	N	Mean
TOTAL_INSOURCED_COST	40	135708
PRED_IN_COST_NEW	39	135799.25
PRED_IN_COST_LOG	39	11.2696393
ERROR1_ACT	39	-7.0213513

Variable	Std Dev	Minimum	Maximum
TOTAL_INSOURCED_COST	150068.67	0	691200
PRED_IN_COST_NEW	172988.02	9474.78	867799.57
PRED_IN_COST_LOG	1.0980889	9.1563888	13.6737161
ERROR1_ACT	34.8613776	-103.8299626	72.7442827

Looking at the means of the actual and the predicted in-sourced costs, we see that the log specification of the model does a better job of cost prediction. The mean error is -7.102% which is a small value. We also see that all the predicted costs are constrained in the positive direction as required.

Comparing the predictions and errors of the sub-sample of in-sourced programs given below and the means of the predictions and errors from above, we see that the log specification model provides better cost estimates for the data provided.

### **5.15.2 Comparison of Actual and Predicted Costs for Sub-Sample of Data From the Log Specification Model**

The increased accuracy and effectiveness of the log specification of the cost model can be seen from table 5.44, where a comparison of the predicted costs and associated errors can be made for a sub-sample of the data.

**Table 5.44 Comparison of Actual and Predicted Costs for Sub-Sample of Data From the Log Specification Model**

Organization	TOTAL_IN_COST (\$ (Actual Measured Cost)	PRED COST from log spec (\$)	PRED COST from normal reg (\$)	ERROR from log spec (%)	ERROR from normal reg (%)
QualChoice	84960	84292.11017	118441.2832	0.786123	-39.4083
Memphis Managed Care Corp	333600	205669.8881	216655.5245	38.34835	35.0553
Health Alliance Plan	376800	271550.8577	226399.5965	27.93236	39.91518
IMS Managed Care, Inc.	37440	44010.33494	49318.03133	-17.54897	-31.7255
CareGuide, Inc.	52320	57981.23142	95784.32581	-10.8204	-83.074

**5.15.3 Combined Actual Cost, Combined Predicted Cost and Combined Error Estimate  
Using the Log Specification of the Cost Model**

As in the linear specification, we can sum up the integrated costs for the various DM programs of each firm and provide a total estimate for each organization. Table 5.45 presents the means and other statistics for the costs, while table 5.46 presents the costs calculated for the whole integrated subset. The variable “sum\_actual\_in\_cost\_log” is the summation of the actual in-house costs for each responding health plan, while the second variable. “sum\_pred\_in\_cost\_log” is the total of the predicted in-sourced costs for each of the health plans in the integrated subset of the data. The variable “error\_act” is the prediction error at the health plan level calculated as detailed in section 5.15.

**Table 5.45 Summing Up the Actual and Predicted Costs and the Error Estimate for Each Organization**

The MEANS Procedure

Variable	N	Mean	Std Dev	Minimum	Maximum
SUM_ACTUAL_IN_COST_LOG	16	339,270.00	386,307.20	0	1,507,200.00
SUM_PRED_IN_COST_LOG	15	353,078.06	382,246.54	47,373.90	1,425,801.08
ERROR_ACT	15	-3.5859624	41.549059	-103.829962	70.2153988

**5.15.4 Cost and Error Estimates for the Whole In-Sourced Subset using the Log Specification of the Cost Model**

The whole integrated subset along with the actual and predicted costs obtained from the log specification of the in-sourced cost model and the error, both summed up at the organization level are as reported below in table 5.46.

**Table 5.46 Cost and Error Estimates for the Full In-Sourced Subset From the Log Specification Model**

Organization	Number of DM programs	SUM of Actual TCE costs (\$)	SUM of predicted TCE costs (\$)	ERROR (%)
Healthy Futures, Inc	1	86400	176109.0877	-103.8299626
HealthPartners	2	244800	376669.232	-53.86815032
Partners HealthCare	1	691200	867799.5654	-25.54970564
Miller & Huffman Outcome Architects, LLC	2	108000	133761.615	-23.85334724
IMS Managed Care, Inc.	5	187200	220051.6747	-17.54897152
CareGuide, Inc.	5	261600	289906.1571	-10.82039644
Solucia Inc	5	43200	47373.89726	-9.661799223
QualChoice	5	424800	421460.5509	0.786122679
Health Alliance Plan	4	1507200	1425801.083	5.400671268
Mountain States Home Care	1	77760	71412.70046	8.162679445
Contra Costa Health Plan	1	119040	108339.5013	8.988994183
Florida Health Care Plans	4	839520	734717.5781	12.48361229
Memphis Managed Care Corp	1	333600	205669.8881	38.34835489
Ault International Medical Management, LLC	1	288000	152763.6563	46.95706377
Quality First Healthcare, Inc.	1	216000	64334.73854	70.21539882

### 5.15.5 Comparison of Coefficients From the Regular and Log Specification of the In-Sourced Cost Model

Presented below in table 5.47 is a comparison between the coefficients of the independent transaction cost variables from both the linear and log specification of the models.

**Table 5.47 Coefficient Comparison Between Standard and Log Specification of the Cost Model**

Parameter	DF	Estimate	
		First stage regression	Log specification regression
Intercept	1	277864	12.64238
		(2.15)	(20.00)
TEMPORAL	1	-123102	-1.04877
		(-5.56)	(-9.69)
PHYSICAL	1	-71011	-1.4895
		(-1.06)	(-4.53)
HUMAN	1	142950	1.58807
		(2.97)	(6.75)
CAPITAL	1	-98136	-0.50171
		(-2.81)	(-2.93)
COMPLEXITY	1	18113	0.61511
		(0.4)	(2.78)
SIMILARITY	1	53406	0.54914
		(2.36)	(4.97)
FREQUENCY	1	160158	0.99731
		(5.75)	(7.33)
LAMBDA (inverse mills ratio)	1	-335922	-3.33719
		(-2.41)	(-4.89)

t – Statistics in parenthesis

It can be seen that the log specification of the model preserves the effect of the transaction cost factors on the in-house costs obtained from the linear specification of the integrated cost model, however, as seen from the previous sections, it also produces a better fit for the data ( $R^2 = 0.876$  as compared to 0.6865) and constrains the predicted costs in the positive direction.

### 5.15.6 Comparison of Summed Actual and Predicted Costs From the Regular and Log Specification of the In-Sourced Cost Model

Comparison of the rolled up actual and predicted costs by the normal and the log specification of the model are as shown below in table 5.48. This table provides a side by side representation of the major statistics and means obtained by the normal/linear specification of the cost model with the actual in-house costs recorded. The first row presents the actual costs, while the second row reports the predicted costs obtained via the two separate methods.

**Table 5.48 Comparison of Costs From the Standard and Log Specification of the Cost Model**

Variable	N		STD DEV (\$)		MIN (\$)		MAX (\$)		MEAN (\$)	
	Std Reg	Log Spec	Std Reg	Log Spec	Std Reg	Log Spec	Std Reg	Log Spec	Std Reg	Log Spec
SUM OF ACTUAL COSTS	16	16	386307.2	386307.2	0	0	1507200	1507200	339270	339270
SUM OF PRED COSTS	15	15	322802.5	382246.54	-263296	47373.9	981298.7	1425801.1	361888	353078.06
							Error		9.65%	-3.58%

We see that the log specification model does a good job of TCE cost prediction for the in-sourced subset. Thus, using both the selection and the log specification of the in-house model in tandem, the most appropriate organization form for a particular DM program and the associated costs for the program if it were to be integrated by the health plan can be accurately determined for the consideration of the management and decision makers.

## **Chapter 6. Conclusions and Future Research**

This chapter includes the conclusion to this research and presents directions for further research. Section 6.1 provides the conclusions drawn from the research. In Section 6.2, the use of organizational form and cost analysis and prediction using transaction cost economics as a decision making tool is presented. Section 6.3 provides the directions for future research in this area.

### **6.1 Conclusions**

In the preceding research, we have applied predictive modeling and switching regression techniques to the sourcing decision problem of disease management programs in health plans to determine the factors most heavily influencing this decision and the transaction costs associated with the decision. The results support the hypothesis that transaction cost factors play a major role in determining the organizational form adopted by a health plan for such programs. For the cases that were studied, and from the results of the probit models, one of the principal findings is that while we see that integration becomes more likely as the importance of scheduling increases, temporal specificity is not a significant factor in determining organizational form, as has been found to be the case in other industries. This effect may be due to the fact that while delays in scheduling do have an impact on the transaction costs experienced by the firm, DM programs do not exhibit the phenomena where the delay in one part of the program or task can reverberate and cause delays throughout the rest of the project, as is seen in other industries such as automotive and shipbuilding industries. We also see that the factors for physical asset specificity, human asset specificity, uncertainty and dedicated asset specificity play a vital role in determining the form adopted by the organization. The results provide evidence that integration becomes more probable in the presence of relationship specific physical assets and tools and for capital investment that is specific to that task, service or program.

It is seen that the effect of human asset specificity is the opposite, that is, programs that require more specific skills or knowledge are more likely to be contracted rather than built in-house. In addition, it is seen that organizations tend to outsource programs that are similar to the ones they already have in operation and those that may have a high degree of uncertainty in their performance and effectiveness measurement. The most important findings of this stage are that temporal specificity does not play as major a role as hypothesized in disease management programs, whereas the transaction cost factors physical asset specificity, dedicated asset specificity and uncertainty are major factors in deciding the organizational form chosen by a health plan for a particular disease management program.

Finally, the model to predict the in-sourced costs for DM programs is constructed. The results indicate that the factors for temporal specificity, human asset specificity, dedicated assets and similarity exert their influence on the costs of internal organization, whereas the primary effect of physical asset specificity and uncertainty is predominantly on market costs and not on the costs of internal organization.

## **6.2 Organizational Form and Cost Analysis and Prediction as a Decision Making Tool**

Transaction costs are unique in the fact that unlike other costs incurred such as direct and overhead costs, transaction costs are not recorded or measured. However, as has been shown in this research, the costs are significant, and the efficiency and effectiveness of the disease management program can be greatly improved by minimizing the effect of these factors, while improving the profitability of the organization. The prediction results from TCE models can be used as a decision making tool by management executives of health plans. The decision variables in this approach are the organization form and the predicted costs. Depending on the organization and management of the health plan, the type of disease management program planned and the levels of the various TCE factors, the executives can decide the values for the above mentioned variables and feed that into the models, which would provide them with the organization form to be used in that case and also the in-sourced costs for that program. Thus, a decision can be made whether to integrate or contract the program in order to minimize the transaction costs incurred. For example, a health plan that has contracted one or more of its DM programs may find that the transaction costs can be reduced if the programs were to be integrated. Whereas, the transaction costs for a DM program that has been built in-house may be calculated using our models, and if the associated costs can be reduced by contracting to an external DMO, that option can be

exercised by the health plan. Thus, for a highly competitive industry such as health insurance, reducing transaction costs associated with DM programs can give health plans a competitive edge in the market and improve the management and profitability for this facet of its services considerably. In most cases, the decision makers would want to select the form such that both matches their organizational and management objectives as well as minimizes the transaction costs.

Concluding, this research presents a decision making approach for planning the setup and management procedures for use by planning and management executives in health plans. The importance of internal organization costs distinct from that of market transaction costs suggests that analyses of integration decisions should encompass costs of organizing within as well as in between firms, rather than focusing only on market transaction costs, as prevalent economic theory suggests. Based on the levels of the various factors and the possible impact, the planners can make a choice in selecting the form that best fulfils their objectives and minimizes the costs. This is a general approach and can be applied to any health plan and its DM programs for form and cost prediction.

### **6.3 Future Work**

Even though the sample size on which the models are built is 93 observations, it captures most of the facets of the disease management industry. Hence, the methodology used to study these observations can be extended to much larger sample sizes covering hundreds of health plans and DM programs. Increasing the sample size will potentially further strengthen the predictive capabilities of the form and cost models. The present study was conducted with a small number of observations in a relatively new industry. In addition to the unique features of disease management programs, the tasks and services in this industry are also influenced by government and federal regulations, which can be taken into account by constructing new proxies for these factors. Also, the independent transaction cost variables are imprecise proxies for the variables of true interest, hence there is a need for the refinement of these variables and for new proxies that permit cross firm and most ideally cross industry comparisons of transaction cost factors and costs. Further, since only data on the internal organization costs were accurately measurable, the burden of estimating the internal cost model was heavily dependent on the integrated subset of the total data set.

A decision support software can also be developed for a general purpose use by management executives and planners in health plans and government health agencies. The proposed software would have a graphical user interface with input screens to allow users to feed the transaction cost factor levels for any disease management program. Depending on the input the software will run the two models and provide the ideal organization form and the transaction costs for the in-sourced form of the program for the consideration of the management.

## References

- [1] A Report on the Florida Medicaid Disease Management Program – Historical Perspective, Start-up, Activities, Current Operations, Future Operations and Expectations. 2000. Agency for Health Care Administration.
- [2] Albertson D. 1997. *Is the Disease Management Ship Coming In?* Employee Benefit News, p.29.
- [3] Anderson, E. and Schmittlein, D. 1984. *Integration of the Sales Force: An Empirical Examination*. Rand Journal of Economics, Vol.15, pp.385-389.
- [4] Badaracco, Joseph L., Jr. 1991. *The Knowledge Link: How Firms Compete Through Strategic Alliances*. Boston: Harvard Business School Press.
- [5] Bagga, Sudhir. 2003. *Integrative Medicine-A Practical Application*. IEEE, Vol.8, No.3, pp.137-142.
- [6] Blumberg, Michael R., and Donald F. Blumberg. 1994. *Strategic Directions for Optimum Outsourcing*. The Source, Fall/Winter 1994, pp.8-15.
- [7] Breen, Richard. 1996. *Regression Models: Censored, Sample Selected, or Truncated Data*. Sage University Paper No.111. Thousand Oaks: Sage.
- [8] Buzzel, R. 1983. *Is Vertical Integration Profitable?* Harvard Business Review, Vol.61, pp.92-102.
- [9] Casson, Mark. 1986. *The Role of Vertical Integration in the Shipping Industry*. Journal of Transport Economics and Policy, pp.7-29.
- [10] Casson, Mark. 1987. *The Firm and the Market*, Cambridge: MIT Press.
- [11] Chandler, Margaret K. 1961. *Competition between the Inside and Outside Labour Force for the Work of the Industrial firm*. Proceedings of the Industrial Relations Research Association, pp.334-335.

- [12] Coase, Ronald H. 1937. *The Nature of the Firm*. *Economica* n.s., Vol.4, pp.386-405.
- [13] Coase, Ronald H. 1972. *Industrial Organization: A Proposal for Research, Policy Issues and Research Opportunities in Industrial Organization*, in V.R. Fuchs (ed.), New York: National Bureau for Economic Research).
- [14] Coase, Ronald H. 1988. *The Nature of the Firm: Origin, Meaning, Influence*, *Journal of Law, Economics, and Organization*, pp.3-47.
- [15] Corbett, Michael F. 1995. *Top Ten Reasons Companies Outsource*. The Outsourcing Institute, 1995 Buyer's Guide, pp.29-31.
- [16] D'Aveni, Richard. A. and David J. Ravenscraft. 1994. *Economies of Integration Versus Bureaucracy Costs: Does Vertical Integration Improve Performance?* *Academy of Management Journal*, Vol.5, pp.1167-1206.
- [17] *Disease Management and Medicaid*. 2004. The 11<sup>th</sup> Princeton Conference: Managing Cost and Quality through the Health Care Delivery System.
- [18] Disease Management Association of America. <http://www.dmaa.org>.
- [19] Disease Management Purchasing Consortium. <http://www.dismgmt.com>.
- [20] Einstein, Albert B, Jr, Patricia J. Goldsmith, and David W. Maberry. 2006. *Contracting with a Medicare HMO*. *Cancer Economics*.
- [21] Fitzner, Karen A., Cynthia Gomez, and Heather Chappell. 2005. *DMAA Quality and Research Initiative*. *Disease Management*, Vol.8, No.6, pp.392-396.
- [22] Fox, Heidi. 2005. *Florida Medicaid: Future State of Disease Management*. Disease Management RFI Workshop, Agency for Health Care Administration.
- [23] Fuhr, Johannes, and Thorsten Beckers. 2006. *Vertical Governance between Airlines and Airports- A Transaction Cost Analysis*. Center for Network Industries and Infrastructure.
- [24] Gillette, Bill. 2006. *Disease Management Moments*. *Managed Healthcare Executive*, Vol.16, No.3, pp.32-33.
- [25] Griliches, Zvt. 1957. *Specification Bias in Estimates of Production Function*. *Journal of Farm Economics*, Vol.39, pp.8-20.

- [26] Hanoch, G. 1996. *A Multivariate Model of Labor Supply: Methodology and Estimation*. Princeton, N.J., Princeton University Press.
- [27] Harrigan, Kathryn R. 1983. *Strategies for Vertical Integration*, Lexington, Mass.: Lexington Books.
- [28] Harrigan, Kathryn R., and William H. Newman. 1990. *Bases of Interorganization Cooperation: Propensity, Power, Persistence*. *Journal of Management Studies*, Vol.27, No.4, pp.417-434.
- [29] Healthcare Business, October 1999, p.29.
- [30] Heckman, James J. 1979. *Sample Selection Bias as a Specification Error*. *Econometrica*, Vol.47, No.1, pp.153-161.
- [31] Heckman, James J., and Thomas E. Macurdy. 1980. *A Life Cycle Model of Female Labour Supply*. *The Review of Economic Studies*, Vol.47, No.1, pp.47-74.
- [32] Hennart, Jean-François. 1988. *Upstream Vertical Integration in the Aluminum and Tin Industries*. *Journal of Economic Behavior and Organization*, Vol.9, pp.281-299.
- [33] Hoffman, C., Rice, D., & Sung, H. Y. 1996. *Persons with chronic conditions. Their prevalence and costs*. *JAMA* 276, pp.1473-1479.
- [34] Holtz-Eakin, Douglas. 2004. *An Analysis of the Literature on Disease Management Programs*. Congressional Budget office.
- [35] *Humana DM Program Propelled by Strong CHF Results*. 1996. *Disease Management News*, p.1.
- [36] Joskow, Paul L. 1985. *Vertical Integration and Long Term Contracts: The Case of Coal-Burning Electric Generation Plants*. *Journal of Law, Economics and Organization*, Vol.1, pp.33-79.
- [37] Kissell ,Robert and Morton Glantz. 2003. *Optimal Trading Strategies*, AMACOM.
- [38] Klein, Peter G. 1996. *Economic Calculation and the Limits of Organization*. *The Review of Austrian Economics*, Vol.9, No.2, pp.3-28.
- [39] Leeman, Jennifer, and Barbara Mark. 2006. *The Chronic Care Model Versus Disease Management Programs: A Transaction Cost Analysis Approach*. *Health Care Management Review*, Vol.31, No.1, pp.18-25.

- [40] Lehmann, D. R. and O' Shaughnessy, J. 1974. *Difference in attribute importance for different industrial products*. Journal of Marketing, Vol.38, pp.36-42.
- [41] Lewis, A. 1998. *Outsourcing: Lessons Learned*. Disease Management Purchasing Consortium.
- [42] Lewis, A. 1999. *Outsourcing: Disease Management's Magic Bullet*. Disease Management Purchasing Consortium.
- [43] Managed Care magazine. <http://www.managedcaremag.com>.
- [44] Masten, Scott E. 1993. *Transaction Costs, Mistakes, and Performance: Assessing the Importance of Governance*. Managerial and Decision Economics, Vol.27, No.2, pp.119-129.
- [45] Masten, Scott E. 1984. *The Organization of Production: Evidence from the Aerospace Industry*. Journal of Law and Economics, Vol.27, pp.403-417.
- [46] Masten, Scott E., James W. Meehan, and Edward A. Snyder. 1989. *Vertical Integration in the U.S. Auto Industry: A Note on the Influence of Transaction Specific Assets*. Journal of Economic Behaviour and Organization, Vol.12, pp.265-273.
- [47] Masten, Scott E., James W. Meehan, and Edward A. Snyder. 1991. *The Costs of Organization*. Journal of Law, Economics and Organization, Vol.7, pp.1-25.
- [48] Matheson, M., Anne Wilkins and Daphne Psacharopoulos. 2006. *Realizing the promise of Disease Management—Payer Trends and Opportunities in the United States*. The Boston Consulting Group.
- [49] McFetridge D.G and Smith D. 1989. *The Economics of Vertical Disintegration*, The Fraser Institute, Vancouver.
- [50] Monteverde, Kirk, and David J. Teece. 1982. *Supplier Switching Costs and Vertical Integration in the Automobile Industry*. Bell Journal of Economics, Vol.13, Spring 1982a, pp.206-213.
- [51] Monteverde, Kirk, and David J. Teece. 1982. *Appropriable Rents and Quasi-Vertical Integration*. Journal of Law and Economics, Vol.25, 1982b, pp.321-328.
- [52] Photis M. Panayides and K. Cullinane. 2002. *The vertical disintegration of ship management: choice criteria for third party selection and evaluation*. The Hong Kong Polytechnic University Department of Shipping and Transport Logistics, Maritime Policy & Management, Taylor & Francis Ltd, Vol.29, No.1, pp.45-64.

- [53] Pirrong, Stephen C. 1994. *Contracting Practices in Bulk Shipping Markets: A Transaction Cost Explanation*. Journal of Law and Economics, Vol.36, pp.937-976.
- [54] Prahalad, C., and G. Hamel. 1990. *The Core Competence of the Corporation*. Harvard Business Review, pp.79-91.
- [55] Pruitt, Sheri; Annandale, Steve; Epping-Jordan, JoAnne; Fernández Díaz, Jesús M.; Khan, Mahmud; Kisa, Adnan; Klapow, Joshua; Nuño Solinis, Roberto ; Reddy, Srinath; Wagner, Ed. *Innovative care for Chronic Conditions: Building Blocks for Action*. 2002. WHO Global Report.
- [56] Quilty M, Lewis A. 1997. *Case Studies in Disease Management for Congestive Heart Failure*. Medical Interface, p.97.
- [57] Quilty M, Weiner J, Francis M et al. 1997. *Unleashing the Profit potential of Disease Management: Profile of Leading Strategies and Strategists*. Spectrum Disease Management: Strategic Approaches and Program Analysis, Decision Resources.
- [58] Quinn, James Brian, and Frederick G. Hilmer. 1994. *Strategic Outsourcing*. Sloan Management Review, pp.43-55.
- [59] Renaissance Offers 'Sign & Save' Program. 1997. Disease Management News p.5.
- [60] Shelanski, Howard A., and Peter G. Klein. 1995. *Empirical Research in Transaction Cost Economics: A Review and Assessment*. Journal of Law, Economics & Organization, Vol.11, No.2, pp.335-361.
- [61] Silver, M. 1984. *Enterprise and the Scope of the Firm*, Oxford: Martin Robertson.
- [62] Simon, H. A. 1957. *Models of Man*. Wiley, New York.
- [63] Simon, Herbert. 1957. *A Behavioral Model of Rational Choice*, in *Models of Man, Social and Rational: Mathematical Essays on Rational Human Behavior in a Social Setting*. New York: Wiley.
- [64] Smits, Jeroen. 2003. *Estimating the Heckman two-step procedure to control for selection bias with SPSS*.
- [65] Stigler, George J. 1951. *The Division of Labour Is Limited by the Extent of the Market*. Journal of Political Economy, Vol.59, pp.185-193.

- [66] Straub Detmar Dr. and Soon Ang. 1998. *Production and Transaction Economies and IS Outsourcing: A Study of the U.S. Banking Industry*. MIS Quarterly, Vol.22, No.4, pp.535-552.
- [67] Stuckey, John. 1983. *Vertical Integration and Joint Ventures in the Aluminum Industry*, Cambridge, Mass.: Harvard University Press.
- [68] The Florida Medicaid Disease Management Initiative. 2000. Agency for Health care Administration.
- [69] Theil, H. 1957. *Specification Errors and the Estimation of Economic Relationships*. Revue de l'Institut International de Statistique, Vol.25, pp.41-51.
- [70] Thompson, D., Edelsberg, J., Kinsey, K.L., Oster, G. 1998. *Estimated Economic Costs of obesity to U.S Business*. American Journal of Health Promotion, Vol.13, No.2, pp.120-127.
- [71] Vendors Lead Voters. 1997. *Winners in CSC-DMN Survey*. Disease Management News, p.3.
- [72] Walker, Tracey. 2006. *Disease Management Programs Offered by HMOs*. Managed Healthcare Executive, ABI/INFORM Global, Vol.16, No.4, p.60.
- [73] Wallis, J. and North, C. 1986. *Measuring the Transaction Sector in the American Economy, 1870-1970*. Gallman (ed.), Long Term Factors in American Economic Growth, Chicago, University of Chicago Press, pp.95-148.
- [74] Wheatley, Ben. 2002. *Disease Management: Findings from Leading State programs*. Academy Health, Vol.3, No.3.
- [75] Williamson, Oliver E. 1975. *Markets and Hierarchies: Analysis and Antitrust Implications*. Free Press: New York.
- [76] Williamson, Oliver E. 1979. *Transaction-Cost Economics: The Governance of Contractual Relations*. Journal of Law and Economics, Vol.22, No.2, pp.233-261.
- [77] Williamson, Oliver E. 1981. *The Economics of Organisation: The Transaction Cost Approach*. American Journal of Sociology, Vol.87, No.3.
- [78] Williamson, Oliver E. 1985. *The Economic Institutions of Capitalism*, New York: Free Press.

- [79] Williamson, Oliver E. 1989. *Transaction Cost Economics*, in R. Schmalensee and R. D. Willig (eds.), *Handbook of Industrial Organization*, New York: Elsevier Science Publishers B.V, Vol.1.
- [80] Williamson, Oliver E. 1990. *Comparative Economic Organization: The Analysis of Discrete Structural Alternatives*. University of California, Berkeley, School of Business.
- [81] Williamson, Oliver E. 1993. *Opportunism and Its Critics*. *Managerial and Decision Economics*, Vol.14, No.2.
- [82] Williamson, Oliver E. 1996. *The Mechanisms of Governance*. New York: Oxford University.
- [83] Williamson, Oliver E. 2000. *The New Institutional Economics: Taking Stock, Looking Ahead*. *Journal of Economic Literature*, Vol.38, No.3, pp.595-613.

## **Appendices**

## Appendix A. SAS Code

```
/******IMPORTING *****/  
/******THE DATA SET*****/  
/*******/
```

```
proc import datafile=" C:\Documents and Settings\Inchandav\Desktop\SURVEY.xls"  
out=SURVEY replace;  
run;
```

```
/******"SURVEY" IS RAW DATA  
SET*****/  
/******  
*****/
```

```
data survey1;  
set survey;  
if form = 'In-sourced/Integrated' then DEP = 1;  
else if form = 'Outsourced' then DEP = 0;  
run;
```

```
data survey2 ;  
set survey1;  
if DEP in (0,1);  
run;
```

```
/******GETTING RELEVANT FREQUENCIES***  
*****/
```

```
proc freq data = survey2;  
tables DEP form temporal physical human capital complexity  
similarity frequency uncertainty ORGANIZATION DISEASE/
```

## Appendix A (Continued)

```
norow nocol nopercnt;
```

```
TITLE 'Responding organizations and number of respective responses';
```

```
run;
```

```
/******CHECKING CORRELATION BETWEEN ALL 9 VARIABLES IN FULL DATA  
SET******/
```

```
proc CORR data = survey2;
```

```
VAR DEP temporal physical human capital complexity
```

```
similarity frequency uncertainty ;
```

```
TITLE 'Means and correlations for all variables';
```

```
run;
```

```
/* PICK 80 RANDOM DATA POINTS FOR TRAINING SET*/
```

```
data SURVEY3 (DROP = address email position email telephone nr date name);
```

```
set survey2;
```

```
x = ranuni(4546654);
```

```
run;
```

```
proc sort data = SURVEY3; by x; run;
```

```
data TRAINING VALIDATION;
```

```
set SURVEY3;
```

```
if _n_ <= 80 THEN OUTPUT TRAINING;
```

```
ELSE OUTPUT VALIDATION;
```

```
run;
```

```
DATA TRAINING;
```

```
SET TRAINING;
```

```
TYPE = 'TRAINING';
```

## Appendix A (Continued)

**RUN;**

**PROC FREQ DATA = TRAINING;**

**TABLES DEP;**

**TITLE 'Frequencies for the training set';**

**RUN;**

**DATA VALIDATION;**

**SET VALIDATION;**

**TYPE = 'VALIDATION';**

**ACTUAL\_FORM = DEP;**

**RUN;**

**PROC FREQ DATA = VALIDATION;**

**TABLES ACTUAL\_FORM;**

**TITLE 'Frequencies for the validation set';**

**RUN;**

**DATA VALIDATION (DROP = DEP);**

**SET VALIDATION;**

**RUN;**

**/\* BUILDING THE MODEL WITH THE TRAINING SET\*/**

**proc logistic data=TRAINING descending;**

**model DEP = temporal physical human capital complexity**

**similarity frequency uncertainty/LINK=PROBIT ctable**

**pprob=(0.05 to 1 by 0.05);**

**output out=prob XBETA= g predicted=phat;**

**TITLE 'FIRST STAGE SELECTION MODEL';**

**run;**

## Appendix A (Continued)

```
DATA COMBINED;  
SET TRAINING VALIDATION;  
RUN;
```

```
/* RUNNING THE MODEL WITH THE COMBINED SET*/
```

```
proc logistic data=COMBINED descending;  
  model DEP = temporal physical human capital complexity  
  similarity frequency uncertainty/LINK=PROBIT ;  
  output out=prob2 XBETA= g2 predicted=phat2;  
  TITLE 'FIRST STAGE SELECTION MODEL FOR COMBINED DATA SET';  
  run;
```

```
DATA PROB2;  
SET PROB2;  
IF PHAT2 >= 0.55 THEN PRED_FORM = 1;  
ELSE PRED_FORM = 0;  
RUN;
```

```
/******CHECKING CLASSIFICATION OF TRAINING
```

```
SET******/
```

```
PROC FREQ DATA = PROB2;  
TABLES PRED_FORM*DEP/  
norow nocol nopercnt;  
TITLE 'Classification table for Training set';  
RUN;
```

```
/******CHECKING CLASSIFICATION OF VALIDATION
```

```
SET******/
```

## Appendix A (Continued)

```
PROC FREQ DATA = PROB2;  
TABLES PRED_FORM*ACTUAL_FORM/  
norow nocol nopercnt;  
TITLE 'Classification table for validation set';  
RUN;
```

```
/******STEP – I-b *****/  
/******2ND CASE FOR MODEL *****/  
/******TRAINING VALIDADATION FOR LOGISTIC PROBIT*****/  
/******WITH FACTOR UNCERTAINTY REMOVED*****/
```

```
proc logistic data=TRAINING descending;  
model DEP = temporal physical human capital complexity  
similarity frequency/LINK=PROBIT ctable  
pprob=(0.05 to 1 by 0.05);  
output out=prob_NO_UNCERT XBETA= g_NO_UNCERT predicted=phat_NO_UNCERT;  
TITLE 'FIRST STAGE SELECTION MODEL WITH NO UNCERTAINTY';  
run;
```

```
DATA COMBINED_NO_UNCERT;  
SET TRAINING VALIDATION;  
RUN;
```

```
proc logistic data=COMBINED_NO_UNCERT descending;  
model DEP = temporal physical human capital complexity  
similarity frequency /LINK=PROBIT ;  
output out=prob2_NO_UNCERT XBETA= g2_NO_UNCERT predicted=phat2_NO_UNCERT;  
TITLE 'FIRST STAGE SELECTION MODEL FOR COMBINED DATA SET WITH NO  
UNCERTAINTY';
```

## Appendix A (Continued)

**run;**

```
DATA PROB2_NO_UNCERT;  
SET PROB2_NO_UNCERT;  
IF PHAT2_NO_UNCERT >= 0.55 THEN PRED_FORM = 1;  
ELSE PRED_FORM = 0;  
RUN;
```

**/\*\*\*\*\*\*CHECKING CLASSIFICATION OF TRAINING**

**SET\*\*\*\*\***

```
PROC FREQ DATA = PROB2_NO_UNCERT;  
TABLES PRED_FORM*DEP/  
norow nocol nopercent;  
TITLE 'Classification table for training set';  
RUN;
```

**/\*\*\*\*\*\*CHECKING CLASSIFICATION OF VALIDATION**

**SET\*\*\*\*\***

```
PROC FREQ DATA = PROB2_NO_UNCERT;  
TABLES PRED_FORM*ACTUAL_FORM/  
norow nocol nopercent;  
TITLE 'Classification table for validation set';  
RUN;
```

**/\*\*\*\*\*\*PART II \*\*\*\*\***

**/\*\*\*\*\*\*CALCULATION OF IN-SOURCED AND \*\*\*\*\***

**/\*\*\*\*\*\*OUTSOURCED COSTS \*\*\*\*\***

**/\*\*\*\*\*\* (\$60 HOURLY RATE) \*\*\*\*\***

## Appendix A (Continued)

```
data COMBINED2;  
set PROB2;  
IF ACTUAL_FORM ^= . THEN DEP = ACTUAL_FORM;  
RUN;
```

```
data insourced2;  
set COMBINED2;  
if (dep = 1);
```

```
INSOURCED_ADMIN_COST = STARTUP_TIME_INSOURCED_DAYS*8*60;
```

```
if INSOURCED_ADMIN_COST = . then INSOURCED_ADMIN_COST = 0;
```

```
SEARCH_INFO_COST = SEARCH_INFO_TIME_DAYS*8*60;
```

```
if SEARCH_INFO_COST = . then SEARCH_INFO_COST = 0;
```

```
SUPERVISORY_COST = SUPERVISORY_POLICING_TIME_HOURS_*4*12*60;
```

```
if SUPERVISORY_COST= . then SUPERVISORY_COST = 0;
```

```
TOTAL_INSOURCED_COST =  
INSOURCED_ADMIN_COST+SEARCH_INFO_COST+SUPERVISORY_COST;  
run;
```

```
data outsourced2;  
set COMBINED2;  
if (dep = 0);
```

```
SEARCH_INFO_COST = SEARCH_INFO_TIME_DAYS*8*60;
```

## Appendix A (Continued)

```
if SEARCH_INFO_COST = . then SEARCH_INFO_COST = 0;
```

```
SUPERVISORY_COST = SUPERVISORY_POLICING_TIME_HOURS_*4*12*60;
```

```
if SUPERVISORY_COST= . then SUPERVISORY_COST = 0;
```

```
if LEGAL_COST_OUTSOURCED_DOLLARS= . then
```

```
LEGAL_COST_OUTSOURCED_DOLLARS = 0;
```

```
TOTAL_OUTSOURCED_COST =
```

```
LEGAL_COST_OUTSOURCED_DOLLARS+SEARCH_INFO_COST+SUPERVISORY_COST;
```

```
run;
```

```
/* MEANS FOR IN – SOURCED COSTS*/
```

```
proc means data = insourced2 ;
```

```
var INSOURCED_ADMIN_COST SEARCH_INFO_COST SUPERVISORY_COST
```

```
TOTAL_INSOURCED_COST ;
```

```
TITLE 'MEANS FOR IN - SOURCED COSTS';
```

```
run;
```

```
/* MEANS FOR OUTSOURCED COSTS*/
```

```
proc means data = outsourced2 ;
```

```
var LEGAL_COST_OUTSOURCED_DOLLARS SEARCH_INFO_COST
```

```
SUPERVISORY_COST
```

```
TOTAL_OUTSOURCED_COST;
```

```
TITLE 'MEANS FOR OUT - SOURCED COSTS';
```

```
run;
```

## Appendix A (Continued)

```

/*****PART III *****/
/*****CALCULATION OF THE *****/
/*****THE INVERSE MILLS RATIO *****/
/*****FOR THE HECKMAN 2 STAGE ESTIMATION*****/
/*****TO DERIVE THE IN-SOURCED*****/
/*****COST EQUATION *****/
DATA LAMBDA;
SET INSOURCED2;
IF DEP=1 AND TOTAL_INSOURCED_COST ^= . AND TOTAL_INSOURCED_COST ^= 0
THEN DO;
PDFG2 = PDF('NORMAL',G2);
CDFG2 = CDF('NORMAL',G2);
LAMBDA1 = ((1/sqrt(2*3.141592654))*(exp(-G2*G2*0.5)))/CDF('NORMAL',G2);
LAMBDA2 = (PDFG2/CDFG2);
lambda3=(1/sqrt(2*3.141592654)*exp(-1*g2**2/2))/probnorm(g2);
DELTA1 = -LAMBDA1*G2-LAMBDA1*LAMBDA1;
DELTA2 = -LAMBDA2*G2-LAMBDA2*LAMBDA2;
DELTA3 = -LAMBDA3*G2-LAMBDA3*LAMBDA3;
h1=lambda1**2+lambda1*g2;
h2=lambda2**2+lambda2*g2;
h3=lambda3**2+lambda3*g2;
END;
RUN;

PROC MEANS DATA = LAMBDA ;
VAR LAMBDA1 LAMBDA2 lambda3 DELTA1 DELTA2 DELTA3 H1 H2 H3;
TITLE 'RESULTS FOR THE INVERSE MILLS RATIO AND CONTROL FACTOR DELTA';
RUN;

/* BUILDING THE IN – SOURCED COST MODEL */
```

## Appendix A (Continued)

```
PROC REG DATA=LAMBDA;  
MODEL TOTAL_INSOURCED_COST = temporal physical human capital complexity  
similarity frequency LAMBDA1;  
output out=INSOURCED_PRED predicted=PRED_IN_COST;  
TITLE 'ORGANIZATION COST MODEL FOR IN - SOURCED COSTS';  
RUN;
```

```
DATA INSOURCED_PRED;  
SET INSOURCED_PRED;  
ERROR = ( (TOTAL_INSOURCED_COST -  
PRED_IN_COST)/TOTAL_INSOURCED_COST)* 100;  
RUN;
```

```
/* MEANS FOR PREDICTED AND ACTUAL IN – SOURCED COSTS*/
```

```
PROC MEANS DATA = INSOURCED_PRED;  
VAR TOTAL_INSOURCED_COST PRED_IN_COST ;  
TITLE 'Comparison of Actual and predicted in - sourced costs';  
RUN;
```

```
/******CORRECTING HETEROSKEDASTICITY FOR COST
```

```
MODEL*****
```

```
/******
```

```
PROC REG DATA=LAMBDA;  
MODEL TOTAL_INSOURCED_COST = temporal physical human capital complexity  
similarity frequency LAMBDA1;  
output out=INSOURCED_PRED predicted=PRED_IN_COST residual= RES;  
RUN;
```

```
DATA INSOURCED_PRED;  
SET INSOURCED_PRED;
```

## Appendix A (Continued)

RES2 = RES\*RES;

LAMB= -335922;

N= 39;

**RUN;**

**PROC SQL;**

SELECT SUM(DELTA1) AS DELTAS1

FROM INSOURCED\_PRED;

**QUIT;**

**PROC SQL;**

SELECT SUM(RES2) AS RESS

FROM INSOURCED\_PRED;

**QUIT;**

**DATA** INSOURCED\_PRED;

**SET** INSOURCED\_PRED;

DELTAS1 = -19.4699;

RESS = 2.694E11;

VARC = RESS/N-LAMB\*LAMB\*DELTAS1/N;

SEC = sqrt(VARC);

RHO = sqrt(LAMB\*LAMB/VARC);

**If** (lamb<0) **THEN** RHO = 0-RHO;

C = VARC+LAMB\*LAMB\*DELTA1;

RHOI = sqrt(C);

WGT = 1/RHOI;

**RUN;**

**PROC MEANS DATA** = INSOURCED\_PRED;

**VAR** VARC SEC RHO;

## Appendix A (Continued)

**RUN;**

**PROC REG DATA=INSOURCED\_PRED;**

**MODEL** TOTAL\_INSOURCED\_COST = temporal physical human capital complexity  
similarity frequency LAMBDA1;

**weight** WGT;

**output out=INSOURCED\_PRED\_NEW predicted=PRED\_IN\_COST\_NEW residual=**  
**RES\_NEW;**

**title 'CORRECTING HETEROSKEDASTICITY FOR COST MODEL';**

**RUN;**

```
/******  
/******PART IV *****  
/******CALCULATION OF THE *****  
/******COMBINED COSTS*****  
/******FOR EACH HEALTH PLAN IN THE *****  
/******IN – SOURCED DATA SET*****
```

**DATA** INSOURCED\_PRED\_C;

**SET** INSOURCED\_PRED;

**IF** ORGANIZATION = 'Health Alliance Plan'

**THEN** ORGANIZATION = 'Health Alliance Plan';

**RUN;**

**PROC SQL;**

**CREATE TABLE** SUM\_COSTS **AS SELECT** \*,

SUM(TOTAL\_INSOURCED\_COST) **AS** SUM\_ACTUAL\_IN\_COST,

SUM(PRED\_IN\_COST) **AS** SUM\_PRED\_IN\_COST

**FROM** INSOURCED\_PRED\_C

**GROUP BY** ORGANIZATION;

**quit;**



## Appendix A (Continued)

```
IF TOTAL_INSOURCED_COST^=0;
RUN;
PROC MEANS DATA = SUM_COSTS4;
VAR SUM_ACTUAL_IN_COST SUM_PRED_IN_COST ERROR_ACT ERROR_ABS;
TITLE 'MEANS FOR IN SOURCED ORGANIZATION WITH 1 MISSING REMOVED';
RUN;
```

```
/******  
/******
```

```
proc sort data = SUM_COSTS3;
BY ERROR_ACT;
RUN;
```

```
/******PART V *****/
/******LOG SPECIFICATION*****/
/******OF THE IN - SOURCED*****/
/******COST MODEL FOR BETTER MODEL FIT*****/
/******AND POSITIVE CONSTRAINING*****/
```

```
DATA LAMBDA_LOG;
SET LAMBDA;
TEMPORAL_LOG = log(TEMPORAL);
```

```
physical_LOG = log(physical);
human_LOG = log(human);
capital_LOG = log(capital);
complexity_LOG = log(complexity);
similarity_LOG = log(similarity);
frequency_LOG = log(frequency);
```

## Appendix A (Continued)

```
LAMBDA1_LOG = log(LAMBDA1);
IF TOTAL_INSOURCED_COST ^= 0 THEN TOTAL_INSOURCED_COST_LOG =
LOG(TOTAL_INSOURCED_COST);
RUN;
/* LOG SPECIFICATION OF THE COST MODEL */

PROC REG DATA=LAMBDA_LOG;
MODEL TOTAL_INSOURCED_COST_LOG = temporal physical human capital complexity
similarity frequency LAMBDA1;
output out=INSOURCED_PRED_LOG predicted=PRED_IN_COST_LOG;
RUN;
/*****
/*****HETEROSKEDASTICITY CORRECTION FOR LOG
SPEC*****/
/*****
*****/

PROC REG DATA=LAMBDA_LOG;
MODEL TOTAL_INSOURCED_COST_LOG = temporal physical human capital complexity
similarity frequency LAMBDA1;
output out=INSOURCED_PRED_LOG predicted=PRED_IN_COST_LOG residual= RES;
RUN;
/*****
/*****

DATA INSOURCED_PRED_LOG;
SET INSOURCED_PRED_LOG;
RES2 = RES*RES;
LAMB= -3.33719;
N= 39;
RUN;

PROC SQL;
```

## Appendix A (Continued)

```
SELECT SUM(DELTA1) AS DELTAS1
  FROM INSOURCED_PRED_LOG;
QUIT;
```

```
PROC SQL;
  SELECT SUM(RES2) AS RESS
  FROM INSOURCED_PRED_LOG;
QUIT;
```

```
DATA INSOURCED_PRED_LOG;
SET INSOURCED_PRED_LOG;
DELTAS1 = -19.4699;
RESS = 6.433455;
VARC = RESS/N-LAMB*LAMB*DELTAS1/N;
SEC = sqrt(VARC);
RHO = sqrt(LAMB*LAMB/VARC);
If (lamb<0) THEN RHO = 0-RHO;
C = VARC+LAMB*LAMB*DELTA1;
RHOI = sqrt(C);
WGT = 1/RHOI;
RUN;
```

```
PROC MEANS DATA = INSOURCED_PRED_LOG;
VAR VARC SEC RHO;
RUN;
```

```
PROC REG DATA=INSOURCED_PRED_LOG;
MODEL TOTAL_INSOURCED_COST_LOG = temporal physical human capital complexity
similarity frequency LAMBDA1;
weight WGT;
```

**Appendix A (Continued)**

```
output out=INSOURCED_PRED_NEW_LOG predicted=PRED_IN_COST_NEW_LOG  
residual= RES_NEW_LOG;  
title 'CORRECTING HETEROSKEDASTICITY FOR LOG SPEC OF COST MODEL';  
RUN;
```

```
/*  
/  
/  
/
```

```
DATA INSOURCED_PRED_LOG;  
SET INSOURCED_PRED_LOG;  
PRED_IN_COST_NEW = EXP(PRED_IN_COST_LOG);  
ERROR1_ACT = ((TOTAL_INSOURCED_COST -  
PRED_IN_COST_NEW)/TOTAL_INSOURCED_COST)* 100;  
ERROR1_ABS = (ABS(TOTAL_INSOURCED_COST -  
PRED_IN_COST_NEW)/TOTAL_INSOURCED_COST)* 100;  
RUN;
```

```
PROC MEANS DATA = INSOURCED_PRED_LOG;  
VAR TOTAL_INSOURCED_COST PRED_IN_COST_NEW PRED_IN_COST_LOG  
ERROR1_ACT ERROR1_ABS;  
TITLE 'MEANS FOR WHOLE IN SOURCED ORGANIZATION SET - LOG SPEC';  
RUN;
```

```
/*  
/  
/
```

```
DATA INSOURCED_PRED_LOG_B;  
SET INSOURCED_PRED_LOG;  
IF TOTAL_INSOURCED_COST^= 0;  
RUN;  
PROC MEANS DATA = INSOURCED_PRED_LOG_B;
```

**Appendix A (Continued)**

```
VAR TOTAL_INSOURCED_COST PRED_IN_COST_NEW PRED_IN_COST_LOG  
ERROR1_ACT ERROR1_ABS;  
TITLE 'MEANS FOR IN SOURCED ORGANIZATION SET - LOG SPEC WITH 1 MISSING  
REMOVED';  
RUN;
```

```
/*  
*/
```

```
DATA GOOD_ERROR_LOG;  
SET INSOURCED_PRED_LOG;  
WHERE ERROR1_ABS <= 100;  
RUN;
```

```
PROC MEANS DATA = GOOD_ERROR_LOG;  
VAR ERROR1_ABS ;  
RUN;
```

```
/*CREATING TABLE*/  
/*WITH COMBINED COSTS FOR EACH  
ORGANIZATION*/  
/*FOR COMBINED ERROR  
CALCULATION*/
```

```
DATA INSOURCED_PRED_LOG2;  
SET INSOURCED_PRED_LOG;  
IF ORGANIZATION = 'Health Alliance Plan'  
THEN ORGANIZATION = 'Health Alliance Plan';  
RUN;
```

```
/* FREQUENCIES FOR INTEGRATED SUBSET*/
```

## Appendix A (Continued)

```
PROC FREQ DATA = INSOURCED_PRED_LOG2 ;
TABLES ORGANIZATION DISEASE ;
TITLE 'FREQUENCIES FOR IN - SOURCED SUB - SET';
RUN;

/* CALCULATING TOTAL COST PER HEALTH PLAN*/

PROC SQL;
CREATE TABLE SUM_COSTS_LOG AS SELECT *,
SUM(TOTAL_INSOURCED_COST) AS SUM_ACTUAL_IN_COST_LOG,
SUM(PRED_IN_COST_NEW) AS SUM_PRED_IN_COST_LOG
FROM INSOURCED_PRED_LOG2
GROUP BY ORGANIZATION;
quit;

DATA SUM_COSTS_LOG2;
SET SUM_COSTS_LOG;
BY ORGANIZATION;

IF FIRST.ORGANIZATION;
RUN;

DATA SUM_COSTS_LOG2;
SET SUM_COSTS_LOG2;
ERROR_ACT = ((SUM_ACTUAL_IN_COST_LOG -
SUM_PRED_IN_COST_LOG)/SUM_ACTUAL_IN_COST_LOG)* 100;
ERROR_ABS = (ABS(SUM_ACTUAL_IN_COST_LOG -
SUM_PRED_IN_COST_LOG)/SUM_ACTUAL_IN_COST_LOG)* 100;
RUN;

PROC MEANS DATA = SUM_COSTS_LOG2;
```

## Appendix A (Continued)

```
VAR SUM_ACTUAL_IN_COST_LOG SUM_PRED_IN_COST_LOG ERROR_ACT  
ERROR_ABS;  
TITLE 'MEANS FOR WHOLE ORGANIZATION SET';  
RUN;
```

```
DATA SUM_COSTS_LOG3;  
SET SUM_COSTS_LOG2;  
IF TOTAL_INSOURCED_COST ^= 0;  
RUN;
```

```
PROC MEANS DATA = SUM_COSTS_LOG3;  
VAR SUM_ACTUAL_IN_COST_LOG SUM_PRED_IN_COST_LOG ERROR_ACT  
ERROR_ABS;  
TITLE 'MEANS FOR ORGANIZATION SET WITH MISSING REMOVED';  
RUN;
```

```
proc sort data = SUM_COSTS_LOG2;  
BY ERROR_ACT;  
RUN;
```

**Appendix B. Electronic Survey**



**Disease Management (DM) Outsourcing Survey**

**Please take the time to answer the following questions; your input is greatly appreciated. Please include your contact information so the results may be sent to you.**

Preliminary Information:

**Please enter the name of your organization:**

**Please enter your name:**

**Please enter your position:**

**Please enter your email address:**

**Please enter your telephone number:**

**Please enter your mailing address:**

**Appendix B (Continued)**

---

Disease management Questions:

**Q 1) Please enter the disease for which the program has been implemented:**

- Diabetes
  - Asthma
  - Coronary Artery Disease(CAD)
  - Congestive Heart Failure (CHF)
  - Chronic Obstructive Pulmonary Disease (COPD)
  - Other:
- 

**Q 2) Is this Disease Management program:**

- In-sourced/Integrated ([go to Q3-a next](#))
  - Outsourced ([go to Q3-b next](#))
  - other: Please specify
- 

**Q3-a) What was the approx. time spent (in days) in administrative, facility planning, and other start-up tasks prior to implementation of this in-sourced program? (Go to Q 4 next)**

**Q3-b) What was the approx. legal cost (in \$) involved in bargaining, negotiating and drawing up an appropriate contract for this outsourced DM program?**

**Q 4) What was the approx. time spent (in days) to obtain relevant information in preparation for implementing this program?**

Appendix B (Continued)

---

**Q 5) What is the approx. time spent (in hours) per week for supervisory and managerial tasks for this program?**

---

*Please rate the following on a scale from 1 to 5:*

**Q 6) Scheduling requirements for a particular task or service (such as patient interventions, patient/program effectiveness checks, and risk evaluations) are sometimes critical in a particular program. On the other hand, there is more flexibility regarding the timely completion of tasks and services in other disease management programs. Using the scale, rate how important, in terms of costs, it is to have tasks in this program done on schedule.**

1 - "Not Important"    2    3    4    5 - "Very Important"

---

**Q 7) To what extent are the tools and assets such as the clinical databases and feedback systems, predictive models for patient identification, and monitoring and reporting processes specific to this program? Using the scale below, rate the specificity of the assets required for this program.**

1 - "Relatively Standard"  
 2  
 3 - "Somewhat Specific"  
 4  
 5 - "Very Specific"

**Relatively Standard** - the facilities, assets, etc., used in the program can be easily adapted for use by other industries and other disease management programs.

**Somewhat Specific** - the facilities, assets, etc., used in the program can be easily adapted for use by other disease management programs.

**Very Specific** - the facilities, assets, etc., used in the program cannot be easily adapted for use by others, even other disease management programs.

Appendix B (Continued)

---

**Q 8) To what extent are the skills, knowledge, or experience of the program employees specific to the tasks/services involved in this particular program?**

- 1 - "Relatively standard"  
 2  
 3 - "Somewhat specific"  
 4  
 5 - "Very specific"

**Relatively standard** – the skills, knowledge, and experience of the employees used in the program are comparably valued in applications by other industries and other disease management programs.

**Somewhat specific** – the skills, knowledge, and experience of the employees used in the program are comparably valued in applications by other disease management programs.

**Very specific** - the skills, knowledge, and experience of the employees used in the program would not be comparably valued in applications by others, even other disease management programs.

---

**Q 9) Using the scale below, please rate the investment made in the program in terms of capital, facilities, software and equipment for the setup and monitoring of this specific program, which cannot be used for another program.**

- 1 - "Low"    2    3    4    5 - "Very High"
- 

**Q 10) Please rate the complexity of tasks and services involved in this program using the following scale.**

- 1 - "Fairly Simple"    2    3    4    5 - "Very Complex"
- 

**Q 11) How similar is this program to the other disease management programs offered by the health plan?**

- 1 - "Not Similar"    2    3    4    5 - "Very Similar"
-

**Appendix B (Continued)**

**Q 12) Using the scale below, please rate the frequency of contact with the individuals enrolled in the program.**

1 - "Very Rare"    2    3    4    5 - "Very Frequent"

---

**Q 13) Using the scale below, please rate the difficulty in measuring the outcomes, effectiveness and performance of this program.**

1 - "Easy"    2    3    4    5 - "Very Difficult"

---

USF Disease Management Outsourcing Survey