The Pennsylvania State University

The Graduate School

Harold and Inge Marcus Department of
Industrial and Manufacturing Engineering

COST-EFFECTIVE INTERVENTION– INNOVATIVE STRATEGIES FOR PUBLIC HEALTH CARE

A Dissertation in

Industrial Engineering

by

Renfei Yan

Submitted in Partial Fulfillment
of the Requirements
for the Degree of

Doctor of Philosophy

May 2014

The dissertation of Renfei Yan was reviewed and approved* by the following:

Paul Griffin
Professor of Industrial and Manufacturing Engineering
Department Head Chair
Dissertation Advisor, Committee Co-Chair

Harriet Black Nembhard
Professor of Industrial and Manufacturing Engineering
Committee Co-Chair

Vittal Prabhu
Professor of Industrial and Manufacturing Engineering

Douglas Thomas
Professor of Supply Chain and Information Systems

*Signatures are on file in the Graduate School

ABSTRACT

Limits on health care resources mandate that decision makers design intervention strategies that are cost-effective. In this dissertation, three cost-effective models are developed and applied to address three different health care problems.

The first problem is concerned with the evaluation of children's dental care practices using longitudinal claims data. By linking service mix to the stage of care and applying a Markov chain approach, we examine the feasibility of using recurring patterns of care to evaluate how well patients are integrated into the dental system. The results suggest that privately-insured children are better integrated into dental care than Medicaid-insured children. In addition, transition probabilities estimated from longitudinal claims data provide insight on the quality of a dental program. The model enables practitioners to institute changes in the pattern of delivery, to identify problem areas, and to determine if proposed changes should take place. It also provides a mechanism to estimate the level of government funding needed to support a service goal. We illustrate this by applying the model to estimate the financial cost of implementing a component of the Affordable Care Act for the oral health of children.

The second problem deals with the development of appropriate intervention strategies for Tuberculosis transmission in health care settings. The risk of occupational infection by Mycobacterium Tuberculosis among patients and health care workers has received increased attention. An infection risk model based on biological and physical principles was developed by taking into consideration primary intervention measures. The simulation results show that HIV+ patients should be isolated during admission into the hospital, and that a higher screening frequency for health care workers will significantly reduce the infection cases among patients and health care workers. A distributive model with space segmentation was also studied which suggests that susceptible patients

sitting closer to the infectious sources will have a greater risk of being infected. The study will help hospitals in designing their personalized cost-effective intervention strategy based on their specific situation.

The third problem examines the impact of food deserts on obesity as well as interventions to reduce their significance. We focused on finding the association of food retailers and obesity in US adults. A nonlinear parametric regression was developed using publicly available data at the county level. The model suggests that, in metropolitan areas, obesity rate is positively associated with supercenters and convenience stores and negatively associated with grocery stores and specialty food stores. In non-metropolitan areas, obesity rate is positively associated with supercenters and negatively associated with specialty food stores. We estimated the marginal effect on obesity from the addition of a new food retailer type in a geographic region. This will be useful in identifying regions where interventions based on food retailer type would be most effective. We illustrate several possible applications of the model including: an incentive contract that will elicit a desired level of operator's effort while maximizing the foundation's utility, a proper division of the total subsidies into loans and grants, and a strategic design of food store establishment plan by local government.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

My deepest appreciation goes to my advisor, Prof. Paul Griffin, for giving me continuous support and guidance during the past four years in the most patient way. He always praised my work while encouraged me to walk a step further. I could not imagine a better supervisor and mentor that I would ever have.

I would also like to express the deepest appreciation to Dr. Susan Griffin, who is the collaborator for the dental project. She always gave insightful comments and professional suggestions on my work. Every discussion with her has been illuminating.

I would like to thank the rest of my thesis committee: Prof. Harriet Black Nembhard, Prof. Vittal Prabhu and Prof. Douglas Thomas, for their supervision and insightful comments and their understanding for delaying the defense due to my internship.

Special thanks to my husband, Yufan Zhang. He took care of me in so many ways, meeting and marrying him is the best gift from the PhD life.

Thanks all the friends that I have made in Penn State, who enriched my life with so much fun.

Finally, this work is dedicated to my parents, for their unconditional love.

Chapter 1

Introduction to Cost-Effective Health Care Intervention

1.1 Unbalanced Cost and Quality of U.S. Health Care System

In 2010, the national health expenditures were approximately $2.59 trillion in the United States, which amounts to the entire GDP of France − the worlds' fifth-largest economy (Emanuel, 2012). This spending is approximately 17.9% of U.S. gross domestic product (GDP) (National Center for Health Statistics, 2012) with a per capita expenditure of $8,502. This is significantly greater than all other member countries in the Organization for Economic Co-operation and Development (OECD) (National Center for Health Statistics, 2012). The International Federation of Health Plans, a global insurance trade association including more than 100 insurers in 25 countries, surveyed its members on average prices paid for typical medical services and products. Americans are paying higher prices in 22 out of 23 cases as compared to residents of other developed countries (International Federation of Health Plans, 2013). The problem is further compounded by the large variation in prices across states in the U.S.

Further, spending has been increasing over time across all measures. Total U.S. spending on health care services and supplies grew at a rate of about 2-4% of GDP per decade (Congressional Budget Office, 2012). Nearly half of the spending comes from private sources (e.g. private insurance, out-of-pocket expenditures) with the other half coming from public sources (e.g. Medicaid, Medicare). When comparing spending by funding source, all components have been increasing annually, tough at different rates.

The dramatic increase in health care costs presents a significant challenge to both the federal government and the population in general. Spending on Medicaid and Medicare alone rose to 5.6% in

2011 (Congressional Budget Office, 2012). Health care spending is anticipated to continuously rise over the next 10 years, given the aging of the population, the rising costs of health care, and the implementation of the provisions of the Affordable Care Act (Congressional Budget Office, 2012). Between 1999 and 2009, the average annual growth for out-of-pocket spending was 5% (National Center for Health Statistics, 2012). In 2011, 35% of adults 18–64 years of age who were uninsured did not get or delayed seeking needed medical care, compared with 7% of adults with private coverage and 13% of adults with Medicaid (National Center for Health Statistics, 2012). In 2012, 50% of American families cut back seeking medical care over the previous 12 month period due to cost concerns, and 17% families experienced serious financial problems caused by medical bills, with 11% using up all or most of their savings (Kaiser Family Foundation, 2012).

It is important to note that these rising costs did not lead to better outcomes. For example, the U.S. ranks 37th in 2011 on average life expectancy at birth accroding to the statistics of World Health Organization (WHO, 2013). Further, the U.S. spends more on health care than any of the other OECD country. In 2011, 54% of adults aged 18–64 years who were uninsured did not have usual source of health care in the past 12 months. Further, 10% of adults with private coverage, 12% of adults with Medicaid and 6% of children aged 18 and below also had no usual source of care (National Center for Health Statistics, 2012). When looking at the relationship between cost and quality on a state-by-state basis using Medicare as a proxy of the system, those two factors turn out to be negatively correlated − a higher spending actually associated with a lower quality of care (Baicker and Chandra, 2004).

The Patient Protection and Affordable Care Act (ACA) was signed into law in 2010 to "provide affordable, quality health care for all Americans and reduce the growth in health care spending". It is the most significant government expansion since 1965 when Medicare and Medicaid started to take effect as part of the Social Security Act. ACA provides a combination of measures on health care costs control, public health insurance expansion, and private insurance regulation. According to the most recent CBO projection, it is estimated that the legislation will reduce the number of uninsured residents by 27 million by 2022, assuming the bill's provisions have all taken effect (CBO, 2013).

1.2 Cost-Effective Modeling in Health Care Intervention

As pressure on financial spending for health care programs increases, there is a need to implement efficient interventions from a public health perspective. An ideal system would be one that complies with treatment guidelines, delivers the proper treatment, identifies and prioritizes patients in need, considers proper preventive and therapeutic services, and last but not least, achieves all those goals at an affordable price.

Cost-effective modeling is becoming a common type of approach in public health. It is a method of comparing the resource costs and health outcomes of two or more health intervention alternatives in a quantitative way. Based on the results, decision and policy makers may determine which strategy can best serve their needs and make the best use of available resources. The main challenges for implementing those methods though, are how to measure and value the health outcomes (Marsh et al., 2012).

Cohort Markov models are often employed by health economists, which allocate individuals to particulate state and conduct the simulation at an aggregate level. Such cohort models require that individuals in each state are homogeneous. Other individual-level modeling techniques were proposed as well, including discrete event simulation and agent-based models (Caro, Möller, and Getsios, 2010). Some argue that discrete event simulation is more accurate than cohort Markov models, since the experience of each individual is modeled over time according to the occurred events and their consequences (Caro, Möller, and Getsios, 2010). However, since discrete event simulation requires more data and information than the corresponding Markov model, it can be a challenge for application in some cases.

If decision makers are to value a broader range of outcomes, several alternative methods are often employed, including: cost-consequence analysis (CCA), cost-benefit analysis (CBA), subjective well-being approaches (SWB) and multi-criteria decision analysis (MCDA). The CCA approach lays out all the outcomes with the associated costs, so that policy makers may do the trade-off based on their own judgment. The results of this method depend on the weight that policy makers put on each outcome and can therefore be quite subjective (Marsh et al., 2012). The CBA method assigns a monetary value to both the social cost and benefit so that the outcome measures are directly comparable (Smith, 2000). The challenge with CBA is that the market prices for outcomes are not always available, and in such cases, preference-based methods will need to be employed for the monetary procedure (Marsh et al., 2012). The SWB is an alternative to the preference-based approaches. It measures how an individual's well-being varies as policy changes take place through self-assessment (Dolan et al., 2008). Examples of such assessment include quality-of-life, which includes both adverse and beneficial effects on the individual from the policy change. The SWB method has draw increasing attention in recent years but is still in its infancy (Marsh et al., 2012). The MCDA method constructs a weighted combination of criteria scores for each intervention. It is favored due to its ability of capturing a range of outcomes from various resources (Devlin and Sussex, 2011). The scores or outcomes are collected from decision makers, stakeholders, and researchers. The balancing of those different sources, however, is the key challenge.

In this dissertation, we identify three public health areas where we contribute to the invention policy design by applying some of the cost-effective modeling techniques mentioned above. Those areas are: establishing a quality evaluation scheme in oral health, controlling the spread of Tuberculosis in a clinical setting, and developing a prevention-based intervention strategy for reducing obesity. We briefly describe each of them as follows.

1) Quality Evaluation Scheme

An efficient evaluation scheme shall be designed with the assistance of clinical data based analysis.. Such analysis shall properly identify the efficiency of current clinical procedures, unnecessary elective procedures, disease-related overuse of physician visits and hospitalizations, as well as the fairness of treatment received by different populations (Wennberg et al., 2008). However, not all current evaluation schemes have met all those requirements. For example, evaluation on dental practices focuses on the system utilization rate and neglect the importance of how patients move in the system and the fairness of treatment received by population groups.

The first research problem discussed in chapter 2 is concerned with the evaluation of children's dental care practices using publicly available longitudinal data. The major objective of the study is to examine the feasibility of using patterns of care that emerge from those data to model the appropriateness of care among different insurance groups. Ultimately, the study will facilitate the standardization of quality assessment mechanisms in the future.

2) Hospital-Associated Infection Control

Hospital-associated infections (HAI), or noscomial infections in hospitals, impose significant economic and social consequences on the nation's health care system. Approximately 1.7 million infections occurred in 2002 in U.S. hospitals, and caused or contributed to 99,000 deaths (Klevens et al., 2007). Annual direct hospital costs of treating HAIs range from $35.7 billion to $45 billion, putting great financial burden on hospital and severely reducing the quality-of-life for patients and their families (Scott and Douglas, 2009).

The second research issue discussed in chapter 3 is the consideration of intervention strategies for Tuberculosis (TB) transmission in health care settings. TB can be transmitted by direct or indirect contact, air or combined. While contact transmission of disease forms the majority of HAI cases, transmission through the air is harder to control. New technologies and guidelines have been developed to control the transmission of air-borne diseases. Such devices include mechanical ventilation, high efficiency particulate air (HEPA) filtration, or ultraviolet germicidal irradiation (UVGI) (Eames et al., 2009). It is important to understand the working principle and efficiency of each new technique, as well as the total efficiency when employed together with traditional means of prevention (e.g., surgical masks, national ventilation). Further, administrative control measures such as frequent staff screening and special group isolation (e.g., for HIV patients) will also be considered in the design of the best prevention strategies. While full utilization of all such measures will surely maximize quality, it will also come at the highest cost. We developed a mathematical model for infection risk estimation and applied simulation techniques to determine a individualized and cost-effective intervention strategy for the target health care setting. Our focus is on clinical settings in resource-constrained countries.

3) Preventive Intervention

Early prevention improves quality-of-life in the long-term, and significantly reduces treatment costs. Realizing the importance of prevention, the ACA has invested $15 billion in the Prevention and Public Health Fund to help Americans stop smoking and combat obesity.

Environmental intervention is one important approach for disease prevention. The Healthy Food Financing Initiative, launched by the Obama administration, provides finantial supports for developing stores selling fresh and healthy foods in underserved areas, and aims to eliminate food deserts across the country within seven years. With varieties of food retailer types existing in the market, each type of store will affect the rate of obesity differently. It is therefore important to determine the type and number of stores according to their regional socio-economic conditions, and to strategically design policies and financial incentives.

The last research issue discussed in chapter 4 aims to reduce the impact of food deserts by introducing food retailers into such areas. A nonlinear parametric regression model is developed to quantify the effects for each type of food store in metropolitan and nonmetropolitan counties respectively. The models are then applied to locate retailers in underserved areas in an economic manner. In particular, incentive-based contracts are developed to attract store operators into distressed areas.

Chapter 2

Analyzing Longitudinal Claims Data to Evaluate Children's Dental Care

Abstract

Previous studies divided dental care into four stages: non-users, episodic/emergency users, initial users, and maintenance users. By linking service mix to the stage of care and examining its recurring patterns, our study reveals insights in evaluating the appropriateness of dental practices, especially among Medicaid/SCHIP eligible children. We use data from the Medical Expenditure Panel Survey (MEPS), which recorded dental visits over 2 years, and reported by insurance status. A Markov Chain approach was used to estimate transition probabilities with one year cycle time. The predictive ability of Markov Model is useful in constructing care flow, measuring qualities, identifying deficiencies, and ultimately, establishing a quality assessment mechanism to standardize the evaluation of general dental practices. It may also be applied to estimate the financial cost of implementing the Affordable Care Act for the oral health of children.

2.1 Introduction

Among U.S. children and adolescents, dental caries is one of the most common chronic diseases and the disease most likely to remain untreated (Newacheck et al., 2000). Children from low-income families (< 100% of the federal poverty line (FPL)), although eligible for Medicaid dental insurance, are about twice as likely to have untreated caries than are children in families with incomes above the FPL (Dye, Li, and Thornton-Evans, 2012). Conversely, low-income children are less likely to have utilized dental care within the last year (32.6%) than are high-income children (57.2%)

(Agency for Healthcare Research and Quality, 2012) and are also less likely to have received preventive dental care (Lewis, Robertson, and Phelps, 2005). Lower utilization of dental services can be explained by both demand factors (e.g., lack of insurance or inability to pay) and supply factors (e.g., low Medicaid reimbursements for dental services relative to private market rates).

Changes to the U.S. health care system will likely increase demand for dental care among all U.S. children because pediatric oral health services are included as part of the essential health benefits package (Affordable Care Act §1302) (Department of Health and Human Services, 2013). Thus it will become increasingly important to monitor whether U.S. children, especially those at highest risk for caries, receive timely, quality dental care. A report by the Institutes of Medicine addressing the quality of the U.S. health care system (Institute of Medicine . Committee on Quality of Health Care in America, 2001) defined quality as the extent to which services increase the likelihood of desired health outcomes. It also defined effective health care as not only providing services based on scientific knowledge but also on whether services were delivered to those who would benefit from them, i.e., avoiding underuse and overuse. A recent review of performance measures for dental care sponsored by the National Quality Forum (NQF) found that while process measures for dental care were plenteous, outcome measures were scarce (National Quality Forum, 2012). It further found that among the existing process measures, utilization measures were the most widespread but the NQF panel expressed concerns about the difficulties associated with determining whether patients completed treatment plans and what constituted appropriate levels of utilization.

It is interesting how long these concerns regarding evaluation measures of dental care have existed. A paper published in 1977 noted that the evaluation of dental programs primarily focused on mix of services performed, number of patient visits, and expenditures, which while providing useful information on productivity and utilization, provided little insight into changes occurring in populations served by dental programs (Freed, Marcus, and Forsythe, 1979). This paper and other publications (Drew Ambulatory Review Team (DART), 1976; Marcus, Koch, and Gershen, 1979) proposed using longitudinal data from chart reviews to evaluate the quality of dental care. Based on mix of services provided during the first dental visit, the model identified where patients were initially in the system – non-users, episodic or emergency users, initial users, and maintenance care. A dental program was then evaluated based on how patients moved through the system. For example, a transition among patients in the episodic stage to non-use would be deemed an unsuccessful outcome whereas a transition from initial to maintenance care would be considered a successful outcome.

This work uses dental visiting records from the 2008 and 2009 Medical Expenditure Panel Survey (MEPS) to examine the feasibility of applying this chart review model to publicly available longitudinal data. We compare patient's initial stage of care and transitions among stages of care for three groups of children and adolescents: those with no dental insurance, with private dental insurance, and with Medicaid insurance. We compare our findings to those from the earlier model (Freed, Marcus, and Forsythe, 1979) and also compare our current findings for each class of insurance coverage. Findings from this analysis will provide useful information on the feasibility of using longitudinal claims data to evaluate the quality of dental care provided to children and adolescents as well as the relative quality of care by insurance plan.

2.2 Methodology

2.2.1 Data Description

The Medical Expenditure Panel Survey (MEPS) is a set of large-scale surveys of families and individuals, their medical providers, and employers across the U.S. The household component of MEPS is drawn from a nationally representative subsample of households that participated in the prior year's National Health Interview Survey.

We used data for children and adolescents, aged 1 to 18 years at the end of 2009, from panel 13 of the household component of MEPS. This survey collects data on dental utilization, dental services received, and dental insurance status from families and individuals. Dental service categories in MEPS that are typically delivered to children include: 1) general exam, checkup or consultation; 2) cleaning, prophylaxis, or polish; 3) x-rays, radiographs, or bitewings; 4) fluoride treatment; 5) dental sealants; 6) fillings; 7) crowns; 8) root canals; 9) extractions; 10) abscess or infection treatments; and 11) orthodontics.

Each MEPS panel features five rounds of interviews covering two full calendar years. Figure 2-1 below illustrates the timing and relationship between panels, rounds, and calendar years. The first interview period in Panel 13 began on January 1, 2008, and ended on the date of each reporting unit's Round 1 interview, conducted from March through June 2008. The reference periods for Rounds 2, 3, and 4 varied by household and covered the time between interview dates of the previous round and the current round. The last reference period, Round 5, ended on December 31, 2009. For this analysis, we included respondents reporting the same insurance status for all 5 rounds. Respondents were assigned to an insurance category in each round based on the following criteria: 1) Medicaid if respondent was covered by Medicaid at any time during that round; 2) private if respondent was never covered by Medicaid and had private dental insurance at any time during round; and 3) none if respondent never had insurance.

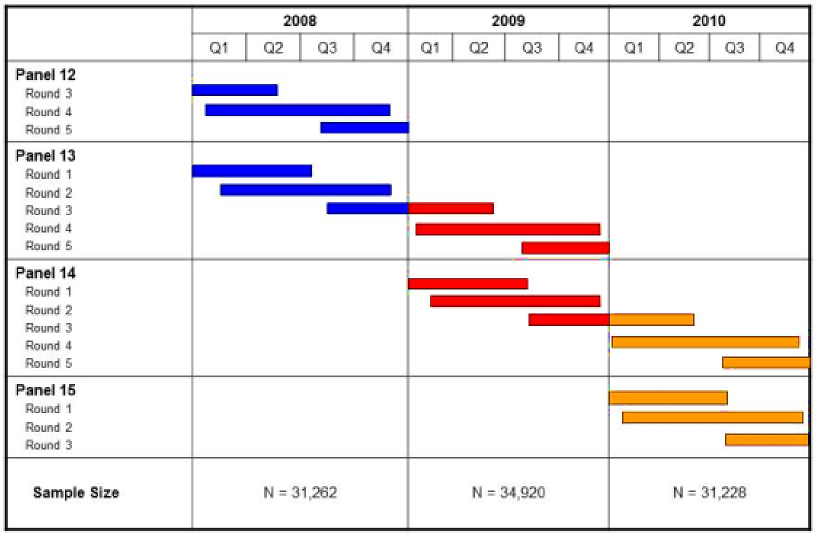


Figure 2-1 MEPS panel design.
Source: MEPS website.

2.2.2 Model

Due to the advantages on the predictive ability and ease of use, a Markov chain is applied to model the pattern of dental care. Previous longitudinal analyses of patients' dental records have identified four mutually exclusive stages of dental care based on services provided (Drew Ambulatory Review Team (DART), 1976). The states of care as described in that study and our criteria for assigning children to each stage are listed below.

1) Episodic (**E**): Patient presents with dental disease that typically causes pain and discomfort and no comprehensive assessment with treatment plan provided. Children in MEPS had to receive at least 1 of the following – crowns, root canals, extractions, or treatment for abscess – and not receive a comprehensive exam.
2) Initial (**I**): Patient receives comprehensive diagnostic evaluation, treatment plan, and treatment for all present dental disease. Children in MEPS had to receive a comprehensive exam and a full set of radiographs.

3) Maintenance (**M**): Patient has completed initial stage and is recalled to assess new disease and to provide preventive services. Children in MEPS received at least one of the following services – periodic exam, preventive care, or fillings – and no comprehensive exam or services defined in **E**.

4) Nonuser (**N**): Patient did not receive any dental service.

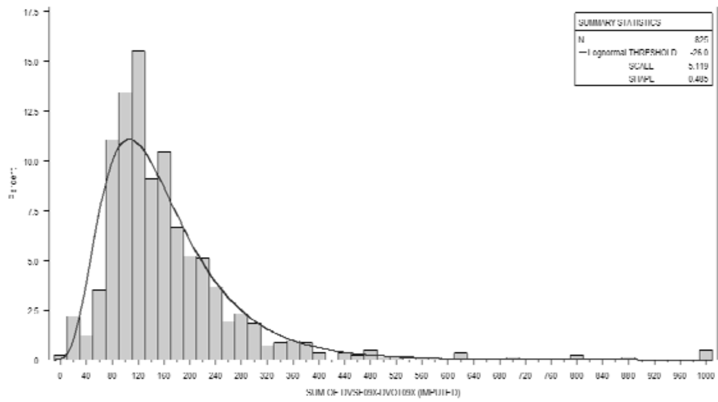
Any visit where orthodontic services were provided was classified as **M**. MEPS data has one code for exam and thus does not distinguish between a comprehensive and periodic exam. A visit was classified as comprehensive if both an exam and radiograph were provided and costs exceeded $140 for privately-insured children, $65 for Medicaid-insured children and $120 for non-insured children.

The cost cut points were set at the median cost for all the visits with both x-rays and exam in 2009 for each insurance type, such that about 13% of visits were comprehensive examinations. This matches a 2006 survey of U.S. dentists, which found that about 13% of examinations were comprehensive (ADA, 2009). Notice in Figure 2-2 that, the cost distribution significantly differ by type of insurance. The selected cut points with the projected proportions of comprehensive examination are presented in Table 2-1.

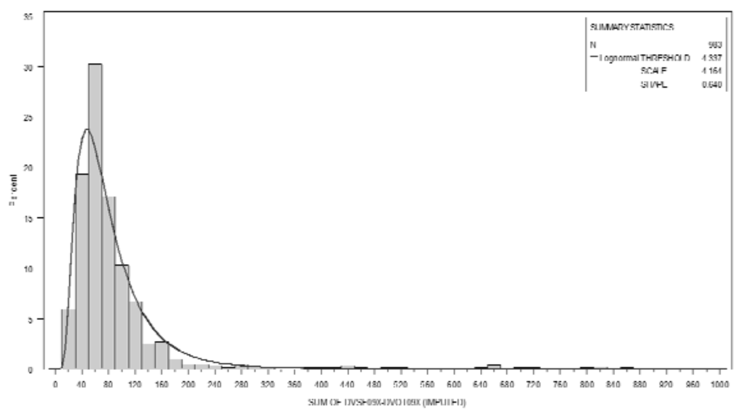
Table 2-1 Cut points on cost and proportion of children defined as "Initial" users (2009)

Insurance	Cutting Point	Proportion of Initial visit
Private	$140	13.0%
Medicaid/SCHIP	$65	12.9%
None	$120	11.6%

a) Private Insured



b) Medicaid Insured



c) Uninsured

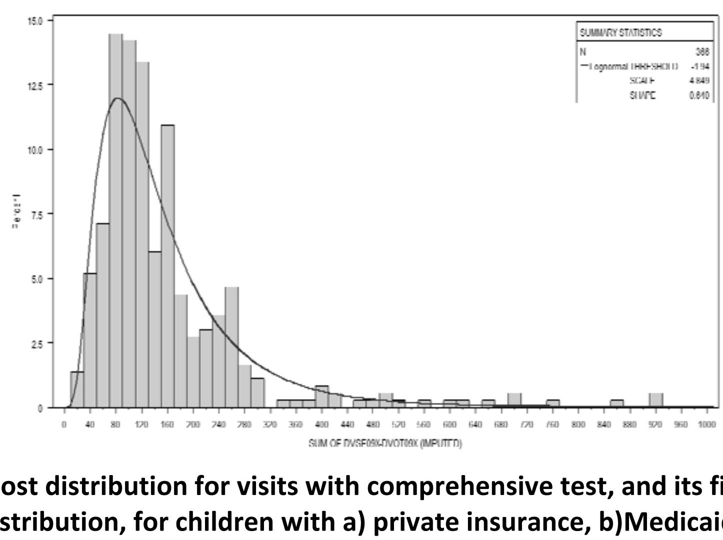


Figure 2-2 The cost distribution for visits with comprehensive test, and its fitting curve to the lognormal distribution, for children with a) private insurance, b)Medicaid/SCHIP and c) noninsurance.

The distribution of services in each stage of care can be used to assess quality of dental care. For example, for patients in M, delivery of preventive interventions should be higher among patient pools deemed to have high caries risk (e.g., Medicaid eligible). In addition, preventive interventions with strong evidence of effectiveness should be more commonly delivered than those with lower levels of evidence.

Transitions across stages of care can also be used to evaluate the quality of care delivered by a provider or provider system (Marcus, Koch, and Gershen, 1979) . In the model as shown in Figure 2-3, N can transit to all stages but M, E can transit to all stages but M, M can transit to all stages but I, I can transit to all stages. To reduce the subjective influence from manipulating the data manually, we modified the model such that transitions between any pairs of stages are possible, as shown in Figure 2-4. We use the transitions among the states of care over time as a means of evaluating the interaction of a population with a care program.

Ideally, nonusers who entered the dental care system would receive a comprehensive exam (I) or receive necessary restorative care (E) and then move on to I. Patients in I could move to E if further treatment was needed or to M otherwise and patients entering the system from E would move to I and then to M. Patients in M would transit back into M. Thus successful transitions would include N to E or I; E to I; I to E; and E, I, or M to M. Conversely, the following transitions would be considered undesirable – any transition to N; N to M (failure to receive comprehensive assessment) and M to E (failure of prevention). Other transitions such as I to I, E to E, and M to I are difficult to evaluate without further information on the patient's risk status. The probability of these transitions should increase with the risk status of the patient pool. In addition, moving from I to I would also occur if a patient changed providers. High probabilities of I to I for low risk patients would be an undesirable outcome for a system in that the system is using resources to provide a comprehensive exam that would not have been necessary if the patient had been sufficiently satisfied with their original provider.

Transitions between cycle 1 (last dental visit during 01/01/2008 – 12/31/2008) and cycle 2 (first dental visit during 01/01/2009 – 12/31/2009) across the different stages of care is modeled as a Markov chain for each insurance status category.

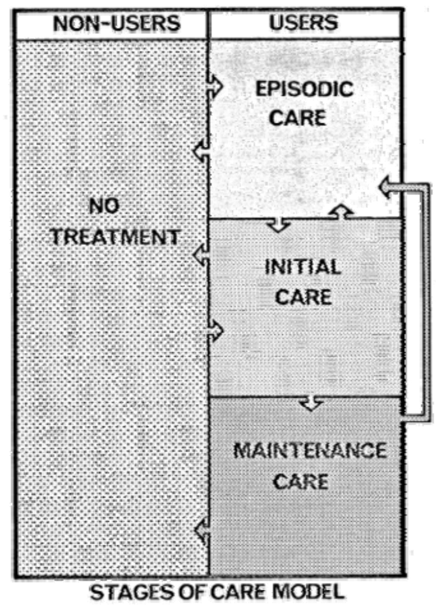


Figure 2-3 Stage of care model proposed by Marcus et al.
Source: (Marcus, Koch, and Gershen, 1979)

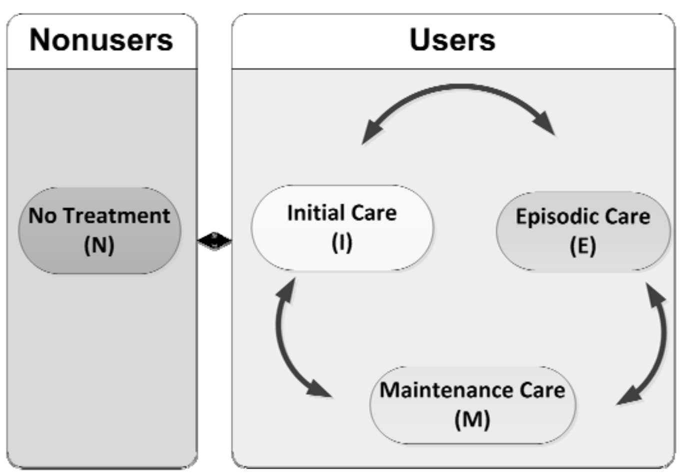


Figure 2-4 Refined model for dental of care.

2.2.3 Analysis

To begin with, exams are conducted on whether a child's insurance status varied by age (1 to 4 years and 5 to 18 years), race/ethnicity (Hispanic, Black, Asian, and White or other) and having had at least one dental visit over the 2-year study period.

We examined services received in each stage of care for each insurance category. For example, for patients in M, delivery of preventive interventions should be higher among patient pools deemed to be at higher risk for dental decay (e.g., Medicaid eligible). In addition, preventive interventions with strong evidence of effectiveness should be more commonly delivered than those with lower levels of evidence.

We then estimated the percentage of children initially in each stage of care and the probability of transitioning across the stages of care for all children and stratified by child's insurance status. For these two analyses, all estimates were standardized by age and race/ethnicity to the weighted distribution of the MEPS sample for this age group, which is representative of the non-institutionalized U.S. population. Two-tailed t-test (alpha=5%) was used to test for significant differences. All statistical analyses were conducted using the SAS software that accounts for the complex multistage design of MEPS.

A comparison was also performed on transition probabilities from this study to those obtained in a study conducted in the 1970s that reviewed dental charts for high-risk children and adults (mean age=21.5 years) attending a university dental clinic that did not provide major orthodontic services (Freed, Marcus, and Forsythe, 1979). For this part of the analysis we did not standardize our transition probabilities by age or race/ethnicity and limited our sample to children, aged 5 years and older. We used three criteria to compare differences in transition probabilities:

1) Absolute difference = Transition probability$_{\text{Freed}}$ – Transition probability$_{\text{MEPS}}$,

2) Relative difference = $\frac{\text{Transition probabilityFreed}}{\text{Transition probabilityMEPS}}$, and

3) Difference in ranking = Ranking of transition probability$_{Freed}$ – Ranking of transition probability$_{MEPS}$. For example, if among transitions out of N, those in Freed were most likely to go to N (rank =1) and those in MEPS were least likely to go to N (rank=4) then the difference in ranking would equal to -3.

Finally, two-way logistic models were fit for each stage of care in 2009, with the main effects of insurance status, stage of care in 2008 (stage 1), race and ethnicity, age, and interaction effect of stage 2008 and insurance status. The reference group are noninsured patients aged 5-18 years who are not Hispanic/Black/Asian and with no dental visit in 2008 (stage 1=N).

2.3 Results

2.3.1 Descriptive Statistics

There were 4,120 children and adolescents (representing 59.4 million nationally) who had the same insurance status for all rounds in Panel 13. Among these children, 24.2% had no dental insurance, 31.3% were insured by Medicaid, and 44.5% had private dental insurance. Children insured by Medicaid were more likely to be younger (less than 5 years old) and Hispanic or Black than were privately insured or uninsured children (Table 2-2). Conversely, privately insured and uninsured children were more likely to be Asian or White and other. Privately insured children were more likely to be younger than uninsured children. Over the 2 year study period, proportion of privately insured children with at least 1 dental visit is 10% higher than either Medicaid-insured or uninsured children (Figure 2-5). This again, proves a higher rate of unmet need among Medicaid children.

Table 2-2 Descriptive statistics by types of insurance.

Characteristic	Insurance					
	None N= 808		Medicaid N= 1949		Private N= 1363	
	%	SE	%	SE	%	SE
1 to 4 years	16.24††**	1.94	25.85	1.4	21.12**	1.36
Hispanic	14.83**	2.00	35.81	2.96	12.05**	1.33
Black	8.17**	1.22	23.11	2.17	10.33**	0.98
Asian	4.9**	1.13	1.16	0.35	6.64**	0.94
Other	72.1**	2.57	39.91	2.55	70.98**	1.86
At least 1 dental visit	58††	2.98	57.44	2.01	69.8**	2.10

* Significantly different from Medicaid at p<0.1
** Significantly different from Medicaid at p<0.05
† Significantly different from privately insured at p<0.1
†† Significantly different from privately insured at p<0.05

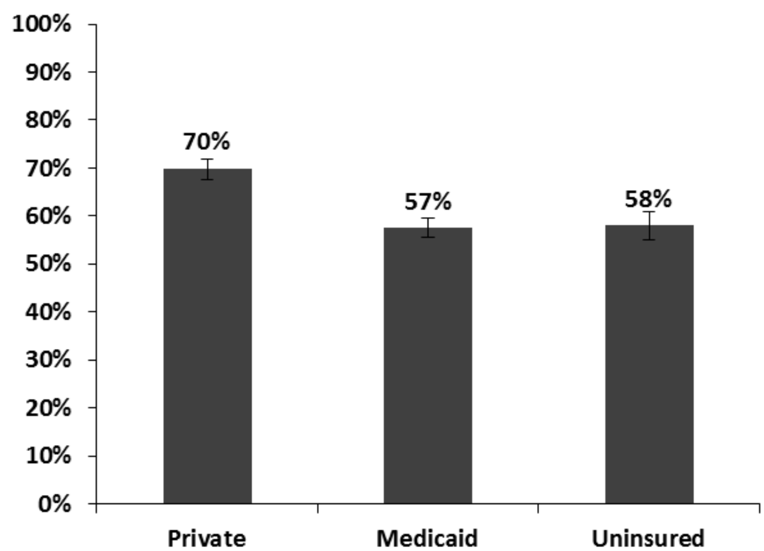
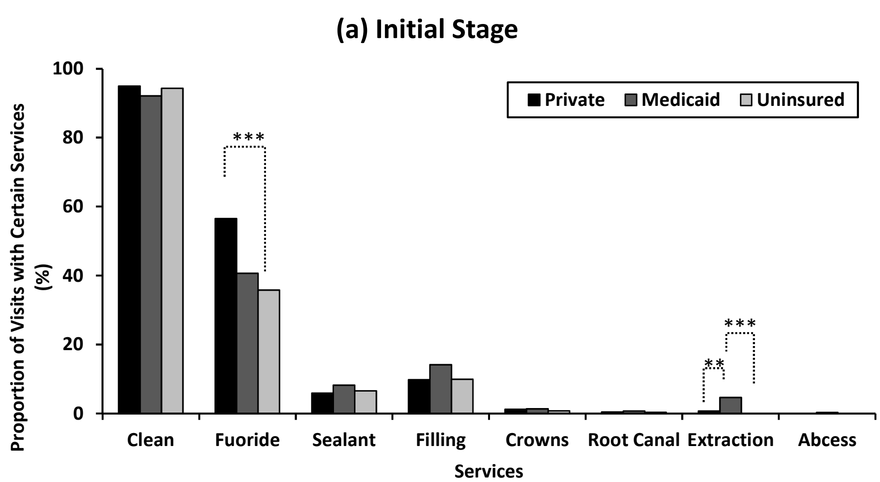


Figure 2-5 Proportion of children with at least one dental visit during 2008-2009.

2.3.2 Distribution of Procedures by Stage of Care

For all visits made during the two years that are of the same stage of care, the distributions of services are shown in Figure 2-6. In each stage of care, the relative receipt of dental services was typically similar across insurance groups. Children in I, all of whom received an exam, were next most likely to receive prophylaxes and topical fluoride than other services. Among children in M, those with Medicaid and private insurance were most likely to receive exams and prophylaxes. The absolute frequency of services, however, did vary by insurance status. Among children in I, Medicaid children were likely to receive extraction care and less likely to receive topical fluoride than were privately insured children. Among children in M, Medicaid-insured children were more likely to receive diagnostic services, topical fluoride and filling, and less likely to receive orthodontic services than were privately insured children. Regardless of insurance status or state of care, children were more likely to receive a prophylaxis than topical fluoride.

Such discrepancies could be due to the variation in risk level and the treatment decisions made by the dentist that are possibly affected by the insurance status of the patients. This result could be used to examine the efficiency of the dental care system and lay out policies to prevent biased treatment decision and to reduce the episodic events ultimately.



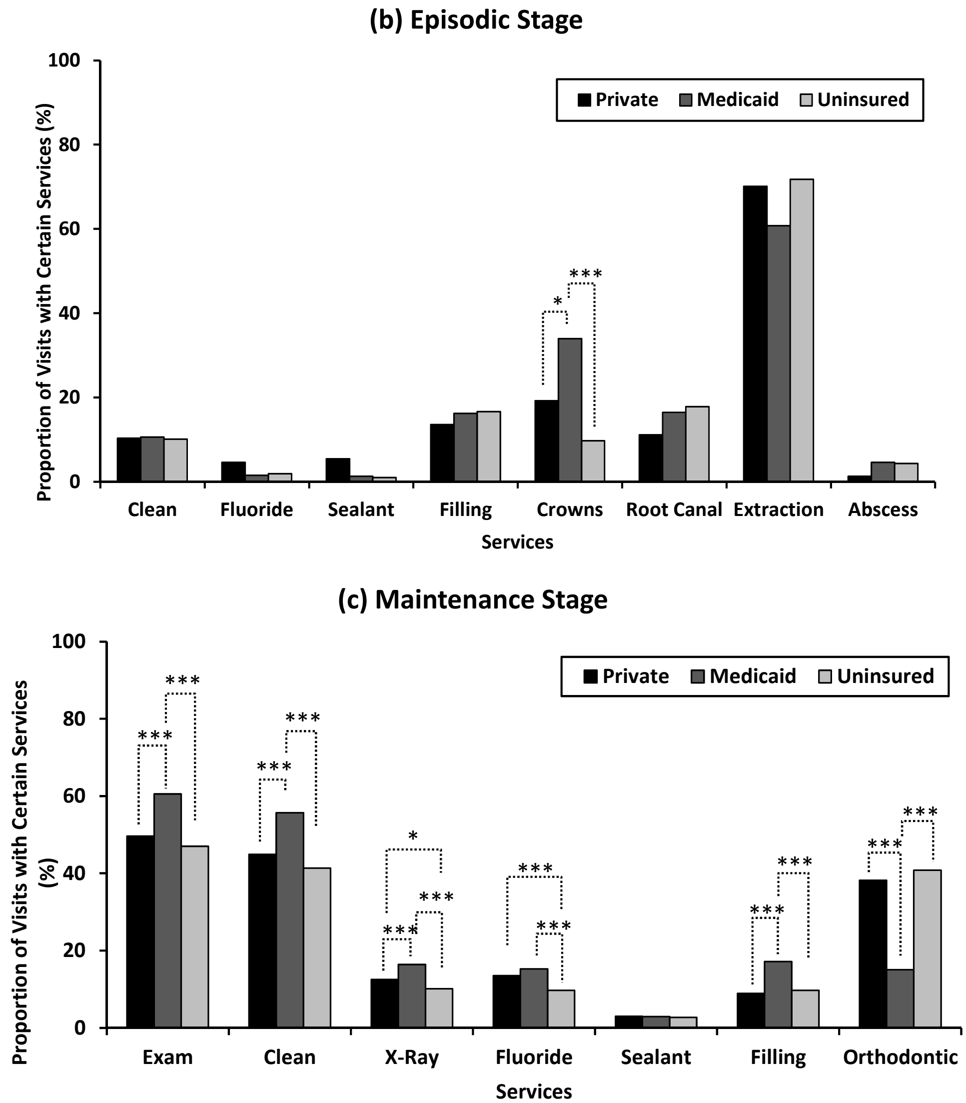


Figure 2-6 Distribution of services by stage of care stratified by insurance status.
Note: significantly different by * p<0.05, ** p<0.01, or *** p<0.001.

2.3.3 Patterns of Care

The first visit in 2008 is used to define the initial state of the patient. Initially, 51.7% of children were in N, 6.9% in I, 1.98% in E, and 39.4% in M (Table 2-3). Privately-insured compared to Medicaid- and non-insured children were less likely to be in N and more likely to be in M. Non-insured children were also more likely to be in M than were Medicaid-insured children. Medicaid-insured children were more likely to be in E than were privately- or non-insured children. Among privately-insured children, the likelihood of being in M vs. N did not differ significantly whereas Medicaid- and non-insured children were more likely to be in N.

Among children initially in N, both non-insured and Medicaid-insured children were more likely to transit back into N than were privately-insured (Table 2-4). Non-insured and Medicaid-insured children were also less likely to transition into M. Among children initially in E, privately-insured children were less likely to transition into N and more likely to transition into M than were Medicaid-insured children. Privately-insured children in E were also more likely to transition into M than were non-insured children. Privately-insured and non-insured children were more likely to transition from I to M than were Medicaid-insured children. Non-insured children in M were less likely to transition to N and more likely to stay in M than were Medicaid-insured children. Privately-insured children were also more likely to stay in M than were Medicaid-insured children.

With the exception of those starting in N, children in our study were most likely to transition to M followed by N whereas in Freed, children were most likely to transition to N and then to M (Table 2-5). The largest absolute differences in transition probabilities between the two studies were for movements into these two stages of care. Among children starting in N, both analyses found that N was the most likely destination. In Freed's study, however, children in N were least likely to transition to M, whereas in MEPS this was the second most frequent destination.

The logistic model coefficients (Coef) and odds ratio (OR) are presented in Table 2-6. The latter is calculated from e^{Coef}, which represent the odds ratio between the corresponding level and the reference level. Weights were included into the model to represents the population in large. While all main and interaction effects were significant in the three models ($P<0.001$), the odds ratios of race were close to 1 with exception of model E. This indicated that race and ethnicity did not have as much influence on transition probability as the other variables.

Table 2-3 Percentage of study population in each stage of care in first time period, standardized for age and race/ethnicity: Medical Panel Expenditure Survey 2008. (Percentage with standard error)

Initial State	No Insurance		Medicaid		Private		All	
	%	SE	%	SE	%	SE	%	SE
N	58.45††	2.17	58.49	2.32	46.2**	1.84	51.72	1.19
E	1.7	0.68	2.66	0.63	1.75	0.40	1.98	0.29
I	4.15**	0.89	11.48	1.66	6.16**	0.86	6.94	0.63
M	35.7**††	2.07	27.38	2.46	45.89**	1.8	39.36	1.24

* Significantly different from Medicaid at p<0.1
** Significantly different from Medicaid at p<0.05
† Significantly different from privately insured at p<0.1
†† Significantly different from privately insured at p<0.05

Table 2-4 Transitional probability standardized for race/ethnicity and age.

Current State	Next State	No Insurance		Medicaid		Private		All	
		%	SE	%	SE	%	SE	%	SE
N	N	76.61††	3.27	71.84	2.99	62.93**	3.15	69.06	1.92
	E	0.37*	0.22	1.85	0.83	1.14	0.66	1.15	0.37
	I	5.00	1.42	6.57	1.64	8.47	1.78	6.68	0.93
	M	18.03††	2.81	19.74	2.75	27.46**	2.74	23.11	1.65
E	N	53.54	15.92	72.22	9.88	28.66**	6.72	47.8	6.12
	E	10.83	9.13	8.88	5.7	5.55	5.28	8.05	3.43
	I	6.51	4.75	1.58	0.95	0.69	0.77	2.29	1.01
	M	29.11††	13.88	17.33	7.68	65.10**	8.49	41.86	6.44
I	N	20.92**††	6.66	50.36	5.04	38.43*	3.96	40.36	3.23
	E	3.61	3.4	4.85	2.75	2.68	1.55	4.08	1.44
	I	20.47†	5.51	14.99	4.07	10.04	3.24	12.77	2.52
	M	55.00**	9.28	29.8	3.71	48.85**	4.67	42.79	3.15
M	N	26.85**	3.71	38.94	5.11	30.62	2.99	31.26	2.05
	E	0.98	0.49	1.28	0.42	1.44	0.48	1.47	0.33
	I	9.37	2.38	9.21	3.12	6.17	1.12	6.71	0.85
	M	62.81**	3.64	50.57	4.53	61.77**	3.14	60.57	2.02

* Significantly different from Medicaid at p<0.1
** Significantly different from Medicaid at p<0.05
† Significantly different from privately insured at p<0.1
†† Significantly different from privately insured at p<0.05

Table 2-5 Comparison between Freed and MEPS analysis.

State		Absolute Difference	Relative Difference	Ranking		
Current	Next			Freed	MEPS	Difference
N	N	26.95	0.41	1	1	0
	I	-1.79	-0.25	2	3	-1
	E	-0.09	-0.07	3	4	-1
	M	-24.77	-0.99	4	2	2
I	N	21.4	0.73	1	2	-1
	I	-4.75	-0.30	3	3	0
	E	-1.9	-0.66	4	4	0
	M	-14.65	-0.28	2	1	1
E	N	17.58	0.46	1	2	-1
	I	5.64	0.59	3	4	-1
	E	6.75	2.45	4	3	1
	M	-30.06	-0.61	2	1	1
M	N	35.99	1.30	1	2	-1
	I	-6.9	-0.83	3	3	0
	E	-1.49	-0.83	4	4	0
	M	-27.6	-0.44	2	1	1

Table 2-6 Logistic regression coefficients and odds ratio by stage of care, 2009.
(N=4120, Weighted Sum= 59,362,810)

		Stage 2							
		N		I		E		M	
Variable	Value	Coef	OR	Coef	OR	Coef	OR	Coef	OR
Intercept		0.96	2.62	-2.62	0.07	-5.18	0.01	-1.4	0.25
Current	I	-2.3	0.1	1	2.71	1.2	3.31	1.87	6.5
	E	-1.31	0.27	0.36	1.44	3.12	22.72	0.89	2.44
	M	-2.25	0.11	0.5	1.65	0.64	1.9	2.11	8.24
Insurance	private	-0.59	0.55	0.58	1.78	0.94	2.56	0.46	1.58
	Medicaid	-0.4	0.67	0.24	1.27	1.3	3.66	0.32	1.38
Stage 1*	I*private	0.64	1.89	-0.74	0.48	-0.39	0.68	-0.46	0.63
Insurance	I*Medicaid	1.27	3.57	0.09	1.09	-0.41	0.67	-1.29	0.27
	E*private	-0.87	0.42	-3.83	0.02	-1.39	0.25	1.3	3.68
	E*Medicaid	1.2	3.33	-1.16	0.31	-1.21	0.3	-1.03	0.36
	M*private	0.74	2.09	-0.82	0.44	-0.38	0.68	-0.52	0.59
	M*Medicaid	1.04	2.83	-0.31	0.73	-0.32	0.73	-0.9	0.41
Age	1-4 Years	0.75	2.12	-0.92	0.4	-1.49	0.23	-0.52	0.6
Race	Hispanic	0.12	1.13	-0.48	0.62	-0.26	0.78	0.05	1.05
	Black	0.39	1.47	-0.74	0.48	-0.29	0.75	-0.16	0.85
	Asian	0.2	1.22	-0.43	0.65	0.63	1.88	-0.12	0.89

Note: Coef stands for logistic model coefficients, and OR stands for odds ratio.

2.4 Discussion

Many of our findings were as expected. Consistent with their higher risk status (Dye, Li, and Thornton-Evans, 2012), Medicaid children who utilized dental care were more likely to receive diagnostic and preventive services than were other children. Medicaid and non-insured children were also more likely to be in N and to stay in N and less likely to move to M than were children with private dental insurance. We also found that among children who utilized the system (starting stage other than N), privately-insured children were more likely to be integrated into care (transition to M) and had fewer unsuccessful outcomes (E to N) than did Medicaid-insured children. One unexpected finding was that among children utilizing the dental system, uninsured children had more successful transitions than did Medicaid-insured children (I to M, and M to M) and fewer unsuccessful outcomes (I to N and M to N). Among users, non-insured children compared to privately-insured children were less likely to transition from E to M and from I to N.

One interesting finding of the analysis was that children, regardless of insurance status, were significantly more likely to receive a prophylaxis than a topical fluoride. There is a strong evidence base for the effectiveness of topical fluoride in preventing caries and thus most guidelines delivering topical fluoride to children at risk for caries (American Dental Association Council on Scientific Affairs, 2006; Hagan, Shaw, and Duncan, 2008). A recent systematic review found insufficient evidence to determine if prophylaxes were effective in preventing dental caries (Azarpazhooh and Main, 2009).

This analysis using MEPS data was similar in some ways to the findings of the earlier Freed's study (Freed, Marcus, and Forsythe, 1979) with the exception that patients in our analysis were more likely to transition to M and less likely to transition to N. This may have been due to differences in study populations – children from all income-levels in our study vs. primarily low-income children and adults in Freed et al. It should also be noted that our study took place almost 40 years after the Freed study and thus factors affecting utilization and treatment protocols may have changed. One notable difference between the studies was the low probability of transitioning from N to M in Freed (<1%) compared to almost 23% in our study. This difference may have been due to our inability to definitively assign patients to I or M because MEPS does not specify whether an examination was comprehensive as would private claims data. It could also, however, have resulted from insurance company (Delta Dental) policy that only allows comprehensive exams for established patients if they are at high risk or have been absent from active treatment for three or more years. Some of the respondents classified as N in this analysis could have been established patients who had not utilized the system within our study time horizon of two years.

Although MEPS data compared to private claims data has less detail about procedures delivered, it does provide some benefits over claims data. Because MEPS data represents a closed system, we can definitively determine that a person has been correctly classified as E or N because we have complete information regarding all potential dental providers whereas for claims data we only know what was done or not done within the provider network. Thus, analysis conducted with MEPS data may be more valid for comparing types of insurers nationally, e.g., Medicaid versus private dental insurers. In addition, transition probabilities for patients with private insurance and no insurance could provide useful benchmarks for the Medicaid dental program. One of the measurement principles set forth in Triple aim is the need for a benchmark or comparison group (Berwick, Nolan, and Whittington, 2008) and indeed at least one previous analysis of the impact of Medicaid dental fees on utilization adopted private utilization as the desired benchmark (Decker, 2011). One limitation of using privately insured children as the benchmark for Medicaid, however, is that children eligible for Medicaid have been shown to be at higher risk for dental disease than children from higher-income families (Dye, Li, and Thornton-Evans, 2012), and thus transition probabilities out of N and into M from previous stages of care may be lower than desired for a higher risk population. On the other hand, because there are multiple determinants of utilization of dental care, some of which may be beyond the control of the dental insurer, transition probabilities for privately insured children may be the most feasible for the Medicaid system.

In conclusion, most current dental performance measures focus on utilization (National Quality Forum, 2012). Examining utilization in the context of transition probabilities across various stages of care allows us to examine how well patients were integrated into the dental care system. For example, we can determine the probability a patient entered the system with high treatment needs (i.e., in E) and whether they eventually transition to M. In addition, transition probabilities across stages of care allow us to estimate the probability a patient stays in the N or M stage for a given period of time and whether new entrants (I or E) eventually transition to M. This analysis suggests that longitudinal claims data, especially beyond 3 years, could provide insight on the quality of a dental program.

2.5 Application to the Affordable Care Act

Based on the transition matrix from Table 2-4, we can calculate the stationary distribution of the patients' stage of care. The results are summarized in Table 2-7. In steady state, Medicaid insured children were most likely to be in the non-use stage and least likely to be in the maintenance stage, while the private insured group had the opposite trend; privately insured children had the lowest chance of being in the non-use stage and the highest chance of being in the maintenance stage. These differences were greater than 15%. This result, once again, shows that the private insured group had greater and more appropriate utilization of the dental care system.

Table 2-7 Steady-state of distribution by type of insurace.

	No Insurance	Medicaid	Private
N	53.0%	60.4%	46.1%
E	1.0%	2.1%	1.5%
I	7.9%	7.9%	7.4%
M	38.1%	29.6%	45.1%

The results can be used to estimate the financial investment required to hit a targeted service level. The Patient Protection and Affordable Care Act of 2010 (as amended by the Healthcare and Education Reconciliation Act of 2010 and referred to collectively as the Affordable Care Act [ACA]) expands dental insurance coverage. The ACA uses Medicaid and CHIP as the way to expand care to those who are unable to afford it, which contains a variety of provisions on oral health (Office of the Legislative Counsel, 2010). In 2014, the percentage of U.S. children without dental insurance will decrease from 25% to almost 0%. In addition, all plans will have to offer select preventive services with no copayments. Thus, implementation of the ACA dental insurance provisions will increase demand for dental services among children who are currently not eligible for Medicaid. Given that the dentist to population ratio has been steadily declining since the 1990s (ADA, 2001), ACA provisions could conceivably reduce the supply of dental care available to current Medicaid eligible children. Our research provides an efficient way to estimate the budget required for implementing those plans, and can assist policymakers to make those adjustments.

Figure 2-7 to 2-11 shows the number of visits by each type and the average cost for each type of visit. As a simple example, if the government desires to increase the number of maintenance (without orthodontic) visits for Medicaid insured children from 0.48 visits per year to the level of private insured children (0.70 visits/year), then the estimated increased annual investment would be as follows,

cost per visit × increased number of visits per person per year × target population

= \$101 × (0.7-0.48) × (58,238,457×31.05%)

= \$410,937,104 / year

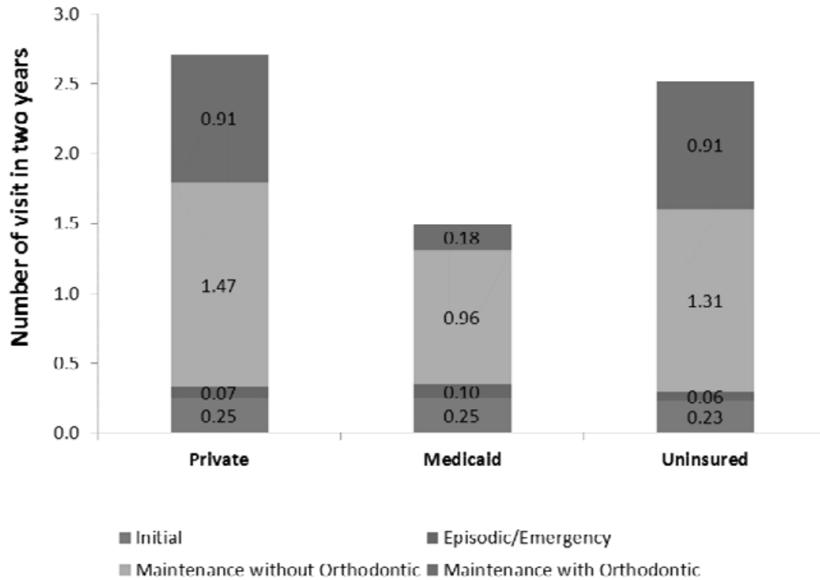


Figure 2-7 Average number of visit during 2008-2009 for all children population.

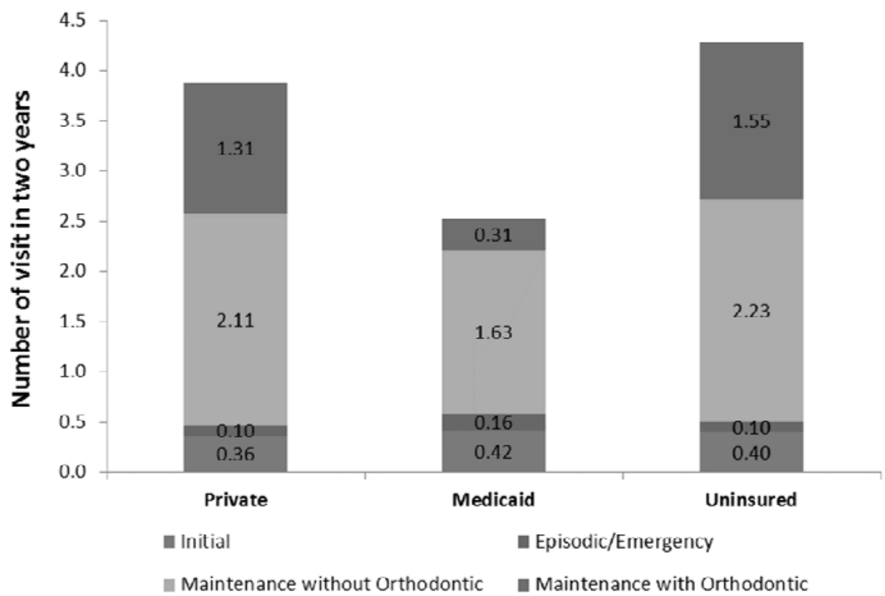


Figure 2-8 Average number of visit during 2008 to 2009 for children with at least one dental visit.

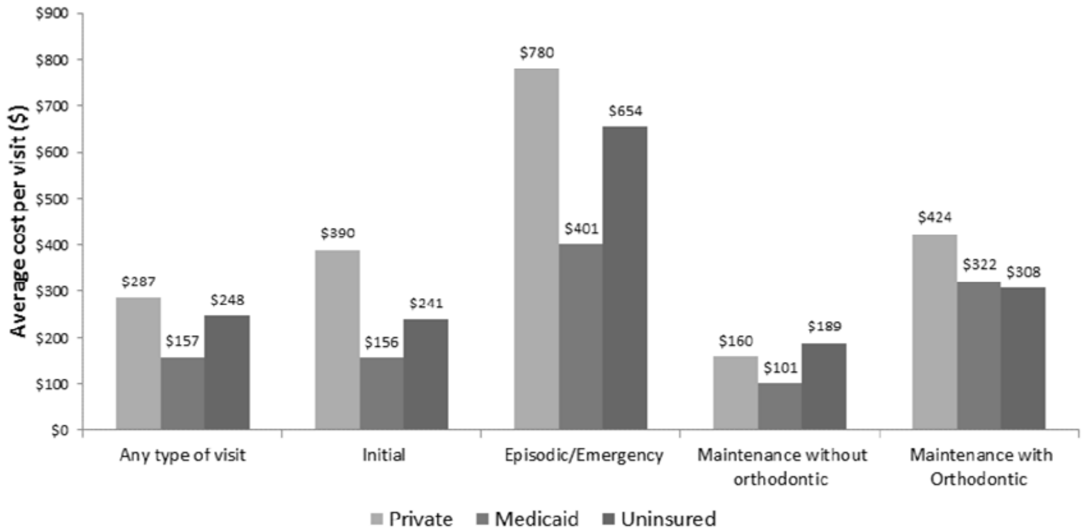


Figure 2-9 Average cost per visit by type of service.

2.6 Conclusions

A study proposed using services provided from chart reviews to determine in which stage of care dental patients started (non-user, emergency/episodic, initial, and maintenance) and to evaluate how well patients were integrated into the system by examining to which stage of care they transitioned and the probability of the transition. We examine the feasibility of applying this model to longitudinal claims data and compare care by patient's insurance status. This analysis suggests privately-insured children are better integrated into dental care than Medicaid-insured and transition probabilities estimated from longitudinal claims data could provide insight on the quality of a dental program. Our findings were similar to the earlier study with the exception that patients in our analysis were more likely to transition to maintenance and less likely to exit the system. The model enables practitioners to institute changes in the pattern of delivery, to identify problem areas, and to determine if proposed changes have taken place. It also provides a mechanism to estimate the level of government funding needed to support a service goal. .

Chapter 3

Intervention Strategy for Occupational Tuberculosis Transmission

Abstract:

The risk of occupational infection by Mycobacterium Tuberculosis (MTB) among patients and health care workers has received increasing attention over the past several years. While multiple control measures and techniques have been developed, there are no personalized guidelines for hospitals to decide which measure should be adopted to manage the infection risk efficiently and economically. In this Chapter, we developed an infection risk model based on biological and physical principles for MTB infection, and took into consideration the primary intervention measures including ventilation, ultraviolet germicidal irradiation, HEPA filtration, surgical masks and respirators. We also considered different isolation policies for HIV patients, secondary transmission from health care workers, as well as the variation of infection risk with regards to the relative location of the infection source. These scenarios were analyzed extensively through simulation and stochastic modeling approaches. In addition, the cost issues of applying different intervention strategies were considered.

3.1 Introduction

3.1.1 Prevalence and Global Control of Tuberculosis

According to the World Health Organization (WHO) Fact Sheet, approximately one-third of the world's population is infected with Tuberculosis (TB) bacteria. TB is second most deadly diseases due to its single infectious agent. In 2010, 8.8 million people were diseased with TB and 1.4 million among them were died from it. About 60% of the global new TB cases occurred in Asia. Sub-Saharan Africa, however, carried the greatest disease rate with over 270 cases per 100 thousand populations.

The WHO's Stop TB Strategy 2015 (WHO, 2006) was developed to dramatically reduce the prevalence of TB worldwide. As these strategies have been implemented, the estimated number of TB cases has declined. It is reported that the TB death rate in 2010 has dropped 40% as compared to 1990. Our objective in this work is to help contribute to the progress of this strategy by designing and evaluating intervention options for health providers in clinics in resource-constrained countries.

3.1.2 Development of Active Tuberculosis

TB is caused by Mycobacterium tuberculosis (MTB), which usually attacks the lungs and may also attack kidney, spine and brain. TB is transmitted through the air. When a person with lung TB disease coughs, sneezes, spits or speaks, bacteria are released into the air. Adjacent people breathing in only a few germs may become infected (CDC). Not all infected individuals are infectious to the others. The development path of TB is illustrated in Figure 3-1. After being infected, the immune systems of most individuals are able to fight off the bacteria from growing. When an individual is infected but not diseased, he/she is considered to have a latent TB infection (LTBI). They will not feel sick or show any symptoms and are not infectious to others. LTBI is detectable via a tuberculin skin test or a special TB blood test. Among the infected individuals, about 10% will develop TB disease over a lifetime, among which 50% will develop TB in 1 year and 80% within 2 years. Common symptoms of lung TB disease include: cough with sputum and/or blood, weakness, weight loss, fever and night sweats. Those symptoms could be mild in the first few months. But if not diagnosed early, one TB patient can infect 10 to 15 individuals in average over one year. Prevention of TB infection by reducing exposure to MTB is thus the most effective way for TB control. As the risk of developing active TB is highest within the first 2 years of infection, an early intervention strategy that targets individuals with recent infection could also be particularly effective as an epidemic control measure.

TB disease is curable by strictly following a standard six-month course of treatment, which consist of: an initial phase of 2 months treatment on isoniazid, rifampin, pyrazinamide, and either ethambutol or streptomycin; followed by continuation phase of treatment on isoniazid and rifampin for 4 more months (WHO). Such treatments should be provided with sufficient supervision from health workers to ensure treatment adherence. If not treated appropriately, TB bacteria may become resistant to TB drugs.

Multidrug-resistant TB (MDR TB) is TB that is resistant to at least two of the most commonly used anti-TB drugs, i.e. isoniazid and rifampin (CDC). These drugs are considered first-line drugs and are used to treat all persons with TB disease for decades. However, resistance to those medicines is growing and is documented in every country (WHO). Second-line drugs can be used to treat MDR-TB, but such drugs can be scarce especially in some resource-constrained countries.

For individuals with human immunodeficiency virus (HIV), their risk of progress to active TB is 21 to 34 times higher than for individuals with normal immune systems, which makes it the strongest risk factor (World Health Organization, 2013). HIV and TB is a deathly combination. As reported by WHO in 2010, almost 350 thousand people died of HIV-associated TB and 25% of HIV patients' deaths are caused by TB. Therefore, HIV patients who are co-infected with TB need special attention during the intervention policy design. The WHO recommended a 12-component approach to integrated TB-HIV services, including actions for prevention and early treatment of infection and disease to prevent rapid progression from LTB infection to active TB disease (WHO, 2006).

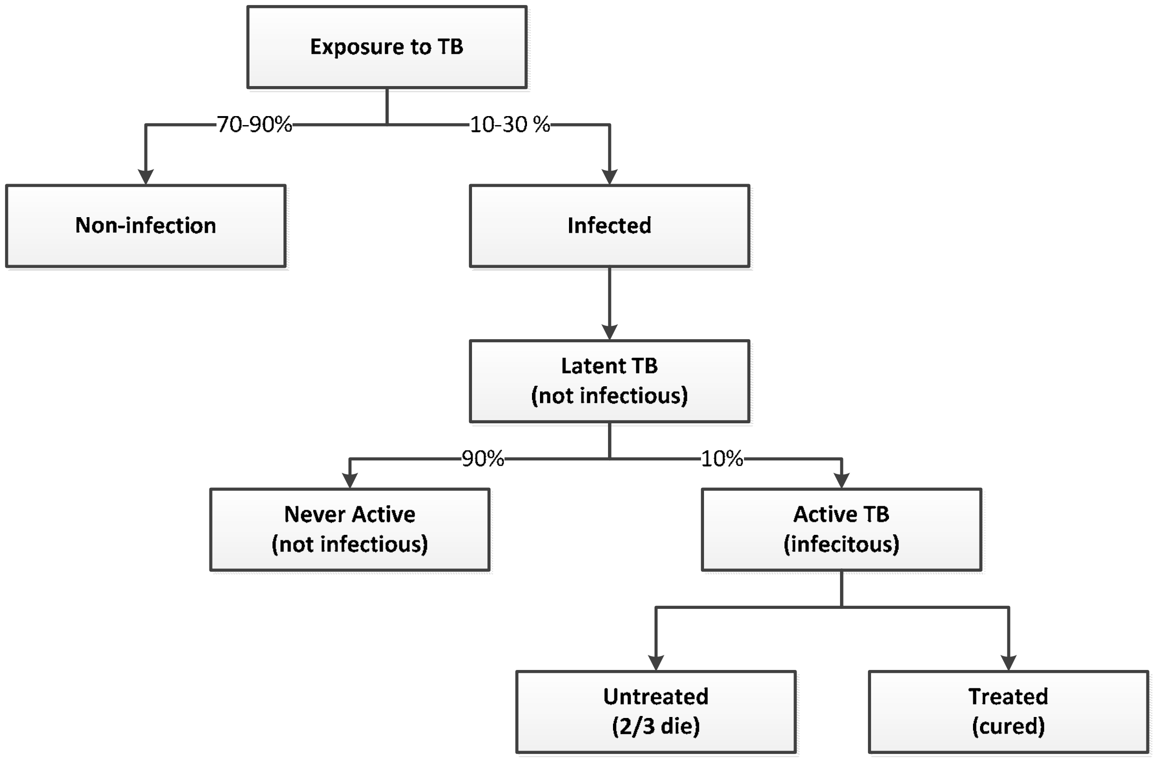


Figure 3-1 Development path of TB disease.

3.1.3 Infection Control in Clinical Settings and Its Challenges

The CDC published *Guidelines for Preventing the Transmission of Mycobacterium tuberculosis in Health-Care Facilities* in 1994 (CDC, 1994) as a response to the prevalence of TB and the non-effective control strategy. According to these guidelines, all health care settings should have an infection control program designed with an emphasis on prompt detection, airborne precautions, and treatment of persons with suspected or confirmed TB disease. Such policies are of particular importance in clinical settings with medium to high risk, where the infection control policies should be developed, reviewed and evaluated periodically.

The wide implementation of the measures has shown success as evidenced by a decrease in the number of TB outbreaks in health care settings and a continuous reduction in hospital-associated transmission of TB to both patients and health care workers (HCW) (CDC, 2005). Following the changes in epidemiology, the CDC updated the TB infection-control guidelines for health care settings (CDC, 2005). The updated guideline emphasized on eliminating the infection risk to HCWs, who are susceptible to patients with unsuspected and undiagnosed infectious TB disease.

The CDC guidelines recommend a three-level hierarchy of control measures based on a risk assessment classification of health care facilities. Those three levels are: administrative, environmental and respiratory-protection control. A general explanation for each of these measures and the accompanying challenges during the implementation in clinical settings is presented below.

A. Administrative Control Measures

Administrative measures is considered as the most important control level as it affects all susceptible individuals who may be exposed to the TB bacteria in the clinical setting. The primary administrative measures include: "conducting a TB risk assessment of the setting, implementing effective work practices for the management of patients with suspected or confirmed TB disease, training and educating HCWs regarding TB, and screening and evaluating HCWs who are at risk for TB disease or infection" (CDC, 2005). Within the screening program, all HCWs with duties for face-to-face contact with suspected or confirmed TB patients should be included. The frequency of screening should be based on the risk classification of the setting.

While more frequent screening could reduce the transmission risk, it will also cause a burden on the hospital and may introduce unnecessary costs. Therefore, a scientific design of the program with an optimal screening frequency is of great importance. In particular, it is important to design the program that works with other intervention strategies to achieve the best effectiveness overall.

B. Environmental Control Measures

The second level of the control program is applying environmental measures to prevent the spread or reduce the concentration of infectious TB bacteria in the air. Primary measures use local exhaust ventilation and general ventilation. Secondary environmental controls manage the airflow to prevent contamination in areas close to the source. Examples include an airborne infection isolation (AII) room, high efficiency particulate air (HEPA) filtration, and ultraviolet germicidal irradiation (UVGI).

Although investment in all of these interventions will achieve the best outcome regarding risk reduction, it also places a financial burden on the health care setting, particularly for those in resource-constrained countries. These decisions should also incorporate the efficacy of each of the available devices.

Moreover, realizing the risk for immunocompromised persons co-infected with TB, those individuals with developed TB disease should be physically separated from other persons to protect both groups. The CDC recommends changing location to avoid exposure for those patients (CDC, 2005).

We will emphasis that specific isolation strategies should be provided based on the special needs of each setting. Considerations for isolation design should at least include the prevalence of HIV patients in the setting, capacity volumes, and other intervention options that have been or will be implemented.

C. Respiratory Protection

The third level of the control program is the use of respiratory protective equipment when there is a high likelihood of exposure to MTB. In certain settings, HCWs might be at risk for both exposure to MTB and mucous membrane exposure to blood borne pathogens. The CDC recommends that HCWs wear a non-fluid-resistant respirator with a full-face shield or the combination product surgical mask/N95 disposable respirator. In certain high-risk locations, such as airborne infection isolation (AII) room, HCWs should wear at least N95 disposable respirators.

While the CDC didn't recommend the use of respiratory protection for patients, it may be cost-effective to ask patients to wear a surgical mask during their visit to reduce their exposure to the MTB. For HCWs, the use of a surgical mask and N95 differ significantly by protection efficiency and cost. Implementation of such equipment needs appropriate consideration and assessment.

3.1.4 Objective

While multiple level control measures have been issued, there are no personalized guidelines for hospitals or clinics to decide which of the activities should be adopted to manage infection risk efficiently and economically. A comprehensive analysis of MTB infection risks under different control strategies within health care settings is needed, including environmental control devices and respiration equipment utilized, screening program designed, as well as isolation strategy implemented, etc. To achieve such goal, a mathematical model for infection risk estimation is required. With such a model, a personalized and cost-effective intervention strategy can be established based on the specific conditions of the health care setting.

3.3 Methodology

3.3.1 Infection Risk Model

In this section, an infection risk model is developed based on biological and physical principles for MTB infection. A simple model for airborne infection is introduced, followed by modifications that consider primary intervention measures such as ventilation, ultraviolet germicidal irradiation, HEPA filtration and masks.

3.3.1.1 Poisson Model

Based on the analysis for a measles outbreak, Riley et al.(Riley, Murphy, and Riley, 1978; Wells and Stone, 1934) established a simple Poisson model for indoor airborne infection that provides a reasonable explanation for observed experimental data. The model includes the probability, *R*, that a susceptible individual breathes in one or more quantum of airborne infection and thus becomes infected,

$$R = 1 - \exp(-D) = 1 - \exp\left(-I \cdot q \cdot p \cdot \frac{t}{Q}\right). \qquad (3.1)$$

A quantum is the minimum number of infectious airborne particles that will initiate an infection. In equation (3.1), I is the number of infectious patients; q is number of quanta produced per infectious patient; p is pulmonary ventilation rate; Q is room ventilation rate with germ-free air. Thus, $I{\cdot}q$ is the total quanta produced per unit time; and $I{\cdot}q/Q$ is the equilibrium concentration of quanta in the air. The cumulative number of quanta (*D*) inhaled during the time of stay would be the concentration $I{\cdot}q/Q$ times the volume of air breathed in $p{\cdot}t$.

The total number of infection cases, *C,* appearing in the next generation is equal to the sum of the individual probabilities of infection over all susceptible patients, *S*. Hence,

$$C = \sum_{i=1}^{S} R_i. \qquad (3.2)$$

The model assumes:

1) A steady state concentration in the room, i.e. a constant release rate of infectious quanta and constant ventilation rate, and the build-up period and biologic delay period was neglected.

2) Complete air mixing, i.e., quanta of infection are evenly dispersed throughout the air of a confined space.
3) Equal susceptibility to a quantum of infection.
4) No difference between individuals.

3.3.1.2 Model Modification

As stated earlier, one major objective of the study is to assess the environmental control measures and respiration protection measures that are most commonly used in clinical settings, i.e., mechanical ventilation (MV), high-efficiency particulate air filtration (HEPA), upper room ultraviolet germicidal irradiation (UVGI), surgical mask and particulate respirator. The efficacy of these measures can be either expressed as an equivalent ventilation rate that alter the value of *Q*, the quanta releasing rate *q*, or the fraction of the cumulative dose that is actually breathed in. In this section, we explain how these measures are incorporated into the model.

a. UVGI

Ko et al. (Ko et al., 2001) proposed to incorporate the disinfection rate in the non-irradiated lower room by upper room UVGI based on a two-zone model. Thus, the efficacy of UVGI can be expressed as an equivalent ventilation rate in the lower room by,

$$\frac{1}{\Delta K_{Lw}} = \frac{1}{\dot{V}/_{V}}\left(\frac{V_L}{V}\right) + \frac{1}{\Delta K_{Uw}}\left(\frac{V}{V_U}\right). \tag{3.3}$$

where $\dot{V}$ is volume of air coming from the upper room per unit of time, *VL* is the volume of the non-irradiated lower room, *VU* is the volume of the irradiated upper room, *V* is the total volume of the room. ΔK_{Lw} and ΔK_{Uw} are the air exchange rates for lower and upper room respectively. As the UVGI is installed in the upper room, ΔK_{Uw} can be expressed as,

$$\Delta K_{Uw} = 3600 \cdot Z \cdot IR, \tag{3.4}$$

where *Z* is the susceptibility of UV and *IR* is the irradiance of UV.

The equivalent increase in ventilation rate due to upper room UVGI, therefore, is defined by,

$$\Delta Q = \Delta K_{Lw} \cdot V_L, \tag{3.5}$$

which alters the denominator of the risk equation (3.1).

b. Mechanical ventilation and HEPA filter

The MV system increases the ventilation rate in unit volume, *A*. HEPA supplementation further increases the effective ventilation by the factor *F*. These two systems alter the denominator of the risk equation (3.1) by,

$$Q' = A \cdot V \cdot F. \tag{3.6}$$

c. Respiratory protection

Surgical masks and particulate respirators have different efficiencies which differ by patient group. In addition, leakage depends on the face-seal. For individuals infected with TB, masks and respirators are used for emitter control. They change the rate that MTB quanta are released. This can be expressed as,

$$q' = q \cdot (1 - E_e^*), \tag{3.7}$$

$$E_e^* = E_e \cdot (1 - L), \tag{3.8}$$

where E_e^* is the effective "efficiency" of emitter control, which as a function of efficiency of the source control (E_e) and leakage rate (L) as expressed in equation (3.8).

For susceptible individuals, the mask and respirator work in a similar way. They alter the effective pulmonary ventilation rate and therefore, adjust the effective number of MTB quanta being inhaled (Gammaitoni and Nucci, 1997). The adjusted pulmonary ventilation rate can be expressed as,

$$p' = p \cdot (1 - E_s^*), \tag{3.9}$$

$$E_s^* = E_s \cdot (1 - L), \tag{3.10}$$

where E_s^* is the effective "Efficiency" of protection, which is as a function of efficiency of protection (E_s) and leakage rate (L) as presented in equation (3.10).

Note that when the same kind of protective equipment is used for individuals that infected with TB and those without, the effective efficiencies of the two groups expressed in equations (3.8) and (3.10) have the same value.

Based on the above discussion for the modifications of the risk model developed by Riley et al., the following model is proposed for the estimation of an individual's cumulative infection risk,

$$R = 1 - \exp(-D) = 1 - \exp\left(-\frac{I \cdot q \cdot p \cdot t \cdot (1 - E_e^*) \cdot (1 - E_s^*)}{A \cdot V \cdot F + \Delta K_{Lw} \cdot (w \cdot l \cdot h_{UV})}\right). \quad (3.11)$$

The variables in the modified risk model are summarized and explained in Table 3-1. The values are either drawn from published studies or based on expert opinion from CDC experts [1].

Table 3-1 Summary of variables in the infection risk model.

Input	Index	Unit	Value	Reference
Breathing Rate	p	m^3/hr	0.34	*Riley et al. 1978*
MTB Quanta Released per Patient	q	qph	2, 13 or 108 with equal possibility	*Nardell, et al. 1993, Kenyon, et al. 1996*
UV Susceptibility	Z	$\mu W^{-1} s^{-1} cm^2$	1.2 to 3	*Xu et al. 2003*
UV Irradiance	IR	$\mu W/cm^2$	partial operation:12; fully operation: 42	*Xu et al. 2003*
Height of UV Installation	h_{uv}	m	2	*CDC expertise*
Air exchange rate in lower room due to UVGI	K_{Lw}	hr^{-1}	Eq. (3.3)	*Riley et al. 1971, Ko et al. 2001*
Height of UV installation	h_{UV}	m	2	*CDC expertise*
Air Exchange Rate by Mechanical Ventilation	A	hr^{-1}	6	*CDC, 2005*
HEPA increased rate of ACH	F	/	3	*Basu et al. 2007*
Efficiency of source control by wearing surgical mask	E_e^m	%	58%	*Nicas et al. 1995*
Efficiency of protection by wearing surgical mask	E_s^m	%	58%	*Nicas et al. 1995*
Efficiency of source control by wearing respirator (N95)	E_e^r	%	95%	*Gammaitoni et al., 1997*
Efficiency of protection by wearing respirator (N95)	E_e^r	%	95%	*Gammaitoni et al., 1997*
leakage rate of surgical mask	L^m	%	20%	*Gammaitoni et al., 1997*
leakage rate of respirator (N95)	L^r	%	10%	*Gammaitoni et al., 1997*

[1]. CDC correspondence: Michele L. Pearson, MD, Division of Healthcare Quality Promotion, Mailstop E-68, Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, GA 30333. mpearson@cdc.gov

3.3.2 Intervention Scenarios

Based on the proposed model, alternative intervention policies can be accessed through simulation. In this section, we will describe each intervention strategy in detail.

3.3.2.1 Hypothesized General Waiting Room

Based on the CDC guidelines (CDC, 2005), a clinical setting is classified as medium risk if three or more TB patients for the preceding year were recorded. We therefore model a hospital with 30 TB patient visits per year, with a general waiting room sized 6×6×3 m^3. The average number of susceptible patients who are not infected with TB before their visit is 30 per day. The actual number of patients, with and without TB disease, visiting each day is assumed to follow a Poisson process. The length of a visitor's stay follows a gamma distribution, which arises naturally in processes for which the waiting times between Poisson distributed events are relevant. The variables and their values for the hypothesized waiting room are summarized in Table 3-2.

Table 3-2 Summary variables for the hypthesized general waiting room.

Parameter	Index	Unit	Value	Reference
# of TB Patient visit per year	I_0	people/day	Poisson (30/365)	*(CDC, 2005)*
Room Width	w	m	6	*CDC Expertise*
Room Length	l	m	6	*CDC Expertise*
Room Height	h	m	3	*CDC Expertise*
# of susceptible patients	S	people/day	Poisson (20)	*CDC Expertise*
Estimated patients exposure time	t	min	Gamma (0.71, 16.94)	*(Ko et al., 2001)*

3.3.2.2 Intervention Devices and Equipment

As discussed in the previous section, several environmental and protection control measures are currently used to reduce TB risks, including mechanical ventilation (MV), high-efficiency particulate air filtration (HEPA), upper room ultraviolet germicidal irradiation (UVGI), surgical mask and particulate respirators. Although the combined use of all the interventions will reduce infection risk the most, each intervention has an associated cost. We therefore focused on finding the most cost-effective strategy from all combinations. Information on the cost of installation and maintenance was obtained from published literature and the manufacturers' website. These include installation cost and annual maintenance cost.

3.3.2.3 Screening Program for HCWs

For administrative control, the CDC (CDC, 2005) recommends that hospital settings adopt a "TB screening program" to test HCWs for infection with MTB. Specifically, the CDC recommends screening "all HCWs who have duties that involve face-to-face contact with patients with suspected or confirmed TB disease…" The frequency of screening is based on the risk classification of the setting. For medium risk settings, HCWs should receive screening annually. For settings classified as potential ongoing transmission, the screening needs to be performed every 8–10 weeks.

The objective of this analysis is to evaluate the effectiveness of TB screening programs including the impact of screening frequency on secondary source infection from HCWs. The medium risk hospital setting described previously will be used here. The simulated screening intervals are one year and 10 weeks. Other assumptions were based on recommendations from CDC experts.

The parameters used in our study are summarized in Table 3-3. 30 HCWs were assigned to the waiting room, each of whom works 8 hours per day, 219 days per year.

Table 3-3 Summary of variables related to the HCWs screening program

Parameter	Index	Unit	Value	Reference
# of HCW working in the waiting room	hcw	/	30	*CDC Expertise*
average working days per year at the waiting room	day_{hcw}	days/year	219	*CDC Expertise*
average working hours per day at the waiting room	t_{hcw}	hours/day	8	*CDC Expertise*
frequency of TB screening for HCW	f_{screen}	days	1 year, 10 weeks	*(CDC, 2005)*

We assumed that at initial state, none of the HCWs are infected or have active TB. Once a HCW is detected to be positive for TB through the screening program, he/she would be removed from their position and replaced by a new HCW. As mentioned previously, infected HCWs are not infectious to others until active TB disease is developed. Studies revealed that after infection, 10% of individuals would develop TB disease over a lifetime, 50% of these individuals develop TB in 1 year following infection, and 80% within 2 years following infection. Therefore, the daily risk of developing TB disease can be estimated. The daily risk of developing TB in the 1st year following infection is approximately 1.4×10^{-4}, and in the 2nd year following infection is approximately 8.8×10^{-5}. Using this information, we may conduct a dynamic simulation on the occupational TB infection. Due to the delay of detection of active TB disease, the diseased HCWs will become the secondary source of infection and transmit MTB to other HCWs and patients. The time of delay depends on the screening program the hospital designed.

3.3.2.4 Isolation Strategies regarding HIV Patients

HIV infection increases the likelihood of progression from LTBI to TB disease. TB disease can also adversely affect the clinical course of HIV infection and AIDS and complicate HIV treatment. Immunocompromised persons, including HIV-infected patients with infectious TB, should be physically separated from other persons to protect both themselves and others. For those patients, changing the location and the scheduling of visits by HCWs is necessary(CDC, 2005). This study focused on the efficiency of risk reduction under different isolation solutions. Four options are considered as shown in Table 3-4. Related variables are summarized in Table 3-5.

Table 3-4 Isolation options for HIV patients.

Isolation Option	Explanation
IS 1	No isolation. All the patients wait at one general waiting area;
IS 2	Isolate HIV+ patients. Only patients with HIV were allocated to the isolation room;
IS 3	Isolate HIV+ and HIV- patients *separately.* All the patients were sent into isolation room, with HIV+ and HIV- patients in different isolation rooms;
IS 4	Isolation HIV+ and HIV- patients *together*. All the patients were sent into isolation room, with HIV+ and HIV- patients mixed together.

Table 3-5 Summarization of variables related to HIV isolation strategy.

Parameter	Index	Unit	Type	Value
# of TB Patients with HIV+	I_{HIV}	people/year	Constant	3
# of susceptible Patients with HIV+	S_{HIV}	people/day	Fit to Poisson distribution	Poisson (1)
Capacity of isolation room	*iso*	people	Constant	5
Isolation Room Width	*ww*	m	Constant	3
Isolation Room Length	*ll*	m	Constant	6
Isolation Room Height	*hh*	m	Constant	3

3.3.3 Distributive Risk Model

So far, we have assumed complete air mixing in the room, i.e., quanta of infection are evenly dispersed throughout the air of a confined space. In this section, we will explore whether there is a variation in infection risk with regard to the relative location to the infection source. A sequential box model was employed (Ko, Thompson, and Nardell, 2004), which divided the space into different segmentations to estimate the infection risk within each segmentation. The model assumes complete air mixing within each box, while movement of air with infectious particles could occur between adjacent boxes.

Three models adopting different number of space segmentations were studied,

1. One-box Model (No segmentation).

This is the same as the model assuming a complete air mixing. All susceptible patients are under the same infection risk calculated by,

$$R = 1 - \exp(-Cpt) \qquad (3.12)$$

where C is the infectious quanta concentration (quanta/m^3), p is the breathing rate (m^3/hr), t is the exposure length of time (hour).

2. Two-box Model.

The whole space is divided equally into two boxes shown as Figure 3-2. The infectious patients will always sit in box 1, while the susceptible patients will sit randomly in the two boxes. In each of the two boxes, the air is completely mixed. The fraction of air transfers between the two boxes per hour is f_i (m^3/hour). The concentration in each box, C_i, can be calculated as

$$\begin{bmatrix} f_1 + A, & -\dfrac{f_2 \cdot V_2}{V_1} \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} = \begin{bmatrix} \dfrac{S_1}{V_1} \\ \dfrac{S_2}{V_2} \end{bmatrix} \qquad (3.13)$$

where A is the air exchange rate per hour, and S_i is the strength of the infectious patient in quanta/hour.

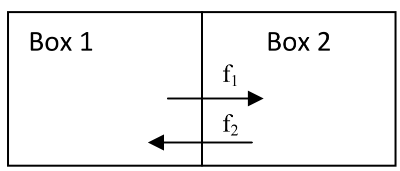


Figure 3-2 Sequential box model for two segmentation case.

3. Four-box Model

This model is similar to the two-box model with equally divided four boxes (Figure 3-3). Note that air transfer only happens between the adjacent two boxes. The calculation of concentration matrix, C_i, is expressed by,

$$\begin{bmatrix} 2f_1+A & -\frac{f_2\cdot V_2}{V_1} & -\frac{f_3\cdot V_3}{V_1} & 0 \\ -\frac{f_1\cdot V_1}{V_2} & 2f_2+A & 0 & -\frac{f_4\cdot V_4}{V_2} \\ -\frac{f_1\cdot V_1}{V_3} & 0 & 2f_3+A & -\frac{f_4\cdot V_4}{V_3} \\ 0 & -\frac{f_2\cdot V_2}{V_4} & -\frac{f_3\cdot V_3}{V_4} & 2f_4+A \end{bmatrix} \begin{bmatrix} C_1 \\ C_2 \\ C_3 \\ C_4 \end{bmatrix} = \begin{bmatrix} \frac{S_1}{V_1} \\ \frac{S_2}{V_2} \\ \frac{S_3}{V_3} \\ \frac{S_4}{V_4} \end{bmatrix} \qquad (3.14)$$

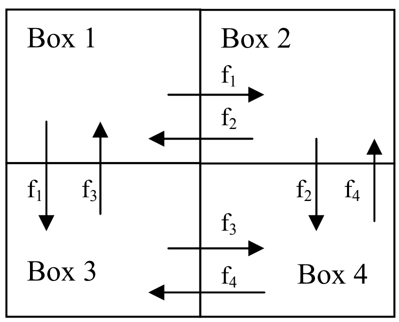


Figure 3-3 Sequential box model for four segmentation case.

3.4 Results

3.4.1 Logic of Simulation

The simulation process for each hospital day is illustrated in Figure 3-4:

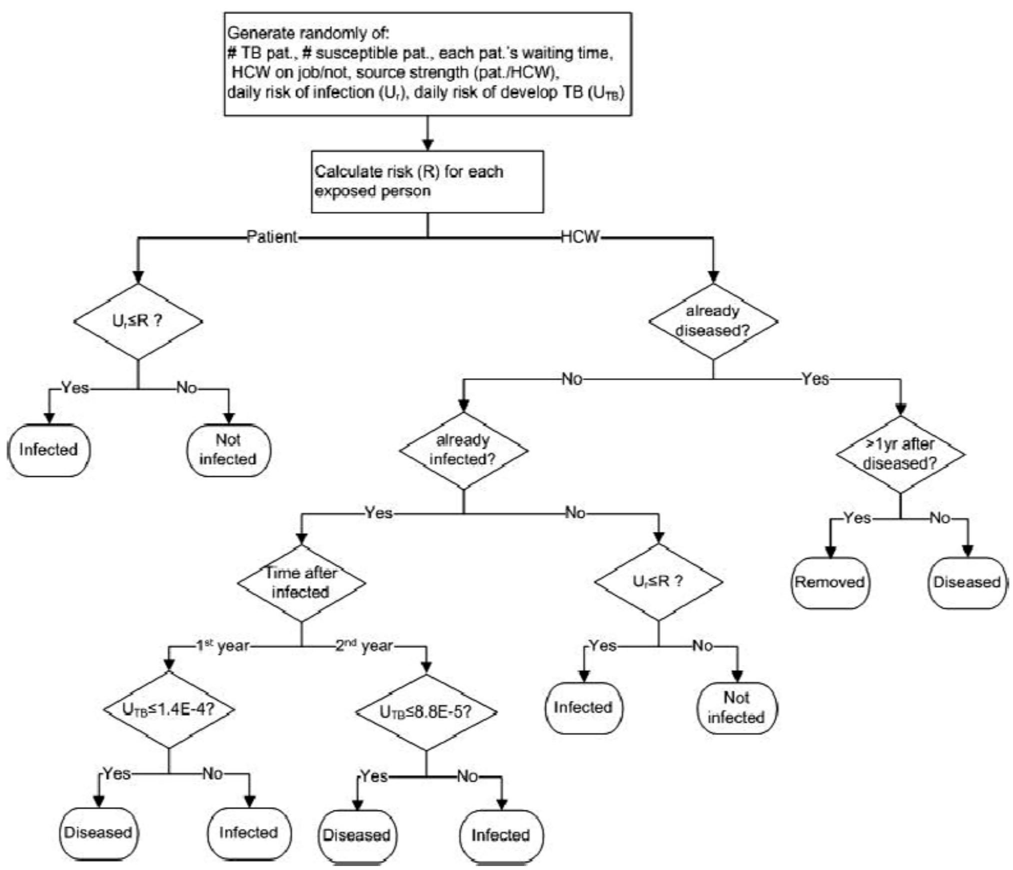
1) For each day, the number of susceptible patients, number of TB patients, waiting time for each patient, and whether or not each HCW is on duty that day is generated based on the appropriate distribution. Both TB patients and the diseased HCWs are potential sources of infection for exposed individuals. Source strength for each of them is approximately 2 qph for chemotherapy-treated TB, 13 qph for active TB, and 108 qph for highly infectious TB (Kenyon et al., 1997; Nardell, 1993).

2) The cumulative daily risk (R) for each patient and HCW is then calculated. A uniformly distributed random value (U_R) between 0 and 1 is then generated. If $U_R \leq R$, the individual is infected. HCWs can only be infected once during the two years of simulation.

3) For infected HCWs, a uniform distributed random value (U_{TB}) between 0 and 1 is generated. During the first year after infection, if $U_{TB} \leq 1.4 \times 10^{-4}$, the HCW will become infected with TB; during the second year, if $U_{TB} \leq 8.8 \times 10^{-5}$, the HCW will then develop TB.

4) For the HCWs who have active TB disease, they will remain on job until the disease is detected by the TB screening program. According to the CDC, the frequency of screening is annual for medium risk settings, and 8-10 weeks for TB transmission ongoing settings.

5) The above steps are repeated each day over a period of two years, and a total of 1,000 iterations was conducted.



,

Figure 3-4 Simulation process for each day.

3.4.2 Preliminary Study on the Proposed Risk Model

A preliminary study that examines all possible combination of the intervention devices and equipment, without adopting any screening program or isolation strategy, was conducted. The primary purpose of this study was to validate the model and eliminate those intervention scenarios that have extremely large infection risk and thus need not to be considered in the later studies. It was assumed that 10 susceptible individuals will visit the clinic following a Poisson process, and that 5 HCWs are assigned to the waiting room each of the 219 working days in a year for 8 hours a day. For each intervention option, Monte Carlo simulation was performed using Matlab (5,000 iterations per scenario). Each iteration constitutes 365 days. The results are presented in Table 3-6 and Figure 3-5. Note that the annual cost is calculated by the sum of installation (discounted over 10 years), maintenance cost per year, and mask and respirator cost per year.

Requiring all HCWs to wear a particulate respirator (N95), which is scenario P9 to P16, reduce HCWs' infection risk significantly as compared to scenarios P1 to P8. Also, MV working at 6 ACH leads to a considerate increase in cost (P1, P4, P9, P12), while UVGI achieves close to the same effectiveness at much lower cost (P3, P6, P11, P14).

Table 3-6 A preliminary study on annual infection cases for patients and HCWs under different intervention options.

Scenario	Intervention Options						Equiva-lent ACH	Annual Infection Cases	
	MV (ACH)	HEPA	UV	surgical mask (patient)	surgical mask (HCW)	N95 (HCW)		Patients	Health care Workers
P1	6				√		6	1.3 ± 0.6	7.4 ± 2.1
P2	3	√			√		9	0.9 ± 0.4	5.1 ± 1.4
P3			√		√		10	0.8 ± 0.4	4.5 ± 1.3
P4	6			√	√		6	0.4 ± 0.2	4.1 ± 1.2
P5	3		√		√		11	0.7 ± 0.3	4.1 ± 1.2
P6			√	√	√		10	0.2 ± 0.1	2.5 ± 0.7
P7	3		√	√	√		11	0.2 ± 0.1	2.3 ± 0.7
P8	3	√		√	√		9	0.2 ± 0.1	2.8 ± 0.8
P9	6					√	6	1.3 ± 0.6	2.2 ± 0.6
P10	3	√				√	9	0.9 ± 0.4	1.5 ± 0.4
P11			√			√	10	0.8 ± 0.3	1.3 ± 0.4
P12	6			√		√	6	0.4 ± 0.2	1.2 ± 0.3
P13	3		√			√	11	0.7 ± 0.3	1.2 ± 0.3
P14			√	√		√	10	0.2 ± 0.1	0.7 ± 0.2
P15	3		√	√		√	11	0.2 ± 0.1	0.6 ± 0.2
P16	3	√		√		√	9	0.2 ± 0.1	0.8 ± 0.2

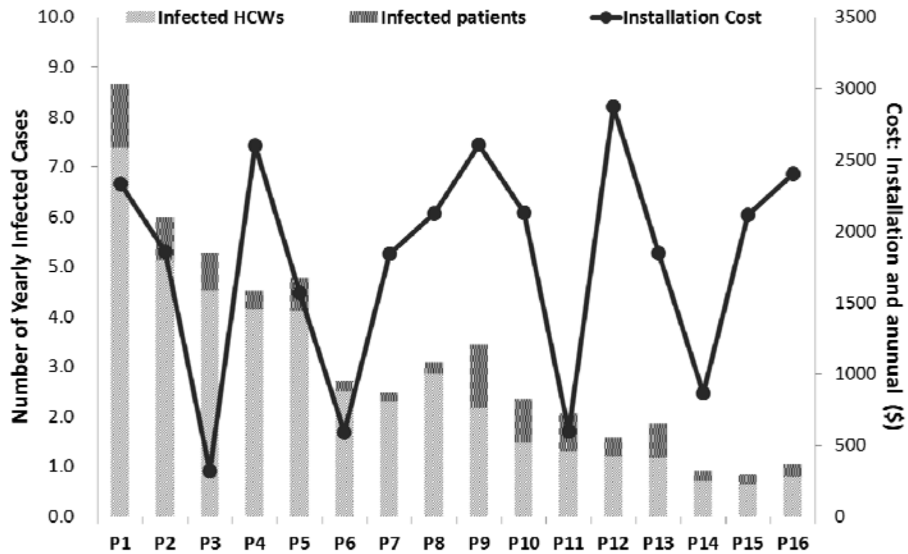


Figure 3-5 Number of infection cases and cost of different intervention options.

Furthermore, assuming that patients and HCWs use surgical masks and N95 respirators respectively, the annual infection cases were calculated at different levels of ACH. The results are presented in Figure 3-6 and Table 3-7. As ACH increases, the number of infections decreases exponentially up to an ACH value of 12, at which point the number of infections levels out. Therefore, an ACH above 12 is not necessary for this specific health care setting.

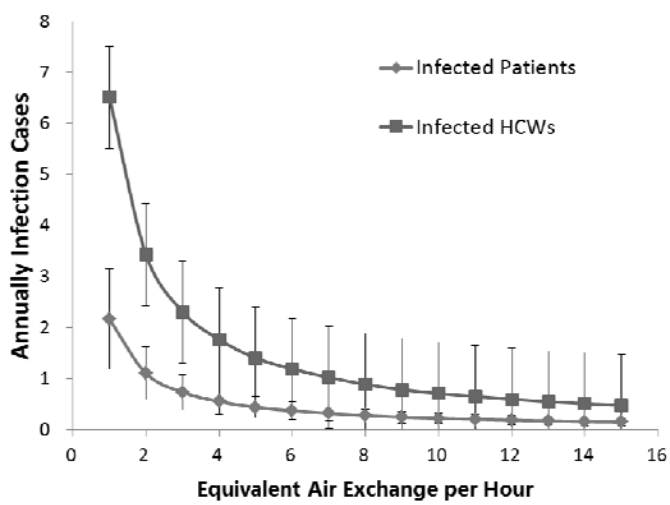


Figure 3-6 Annually Infection Risk vs. Equivalent Air Exchange per Hour. Surgical mask and N95 respirator were used by patients and HCW respectively.

Table 3-7 Annual infection cases under different air exchange rate.

ACH	Hospital Infection per Year					
	Patients			Health care Workers		
1	2.2	±	1.0	6.5	±	1.9
2	1.1	±	0.5	3.4	±	1.0
3	0.7	±	0.3	2.3	±	0.7
4	0.6	±	0.3	1.8	±	0.5
5	0.4	±	0.2	1.4	±	0.4
6	0.4	±	0.2	1.2	±	0.3
7	0.3	±	0.1	1	±	0.3
8	0.3	±	0.1	0.9	±	0.3
9	0.2	±	0.1	0.8	±	0.2
10	0.2	±	0.1	0.7	±	0.2
11	0.2	±	0.1	0.7	±	0.2
12	0.2	±	0.1	0.6	±	0.2
13	0.2	±	0.1	0.6	±	0.2
14	0.2	±	0.1	0.5	±	0.1
15	0.1	±	0.1	0.5	±	0.1

3.4.3 Cost-effectiveness of Intervention Options

As discussed in the previous section, wearing N95 will reduced HCWs' risk of infection significantly. Therefore, we assume that all the HCWs wear a particulate respirator to ensure a low risk of infection. Thus, we narrow the considered control options to six. Table 3-8 shows the control combinations. In addition, a one-year screening interval is adopted.

Table 3-8Combination of the intervention devices and equipment for the cost-effectiveness study.

Control Options	MV (ACH)	HEPA Filtration	UVGI	Surgical Mask Wear by Patients
CO 1	6			
CO 2	3	√		
CO 3	6			√
CO 4	3		√	
CO 5	3		√	√
CO 6	3	√		√

The simulation was conducted for a two year period. The results on an annual basis are presented in Table 3-9. Note that in the second year, the number of cases for patient infection increased. This is due to the presence of more diseased but not detected HCWs who served as a secondary source of infection. The costs for each scenario, which are determined by installation fees, annual maintenance fees and the annual cost for purchasing surgical masks and respirators are presented in Table 3-10. Figure 3-7 shows the results where the cost is based on the sum of "installation cost" depreciated over a 10-year period, annual maintenance fees and the mask/respirator costs; the number of infection cases is the average of the two years. According to this figure, IO5 with MV working at 3 ACH and installed UVGI seems to be the best option. It has the lowest infection risk and third lowest financial cost.

Table 3-9 Infection and diseased cases in two years with application of various control measures.

Control Options	Equivalent ACH	1st year			2nd year		
		Patient Infection	HCW Infection	HCW Diseased	Patient Infection	HCW Infection	HCW Diseased
IO 1	6	3.0 (2.1)	11.7 (4.3)	0.3 (0.6)	4.9 (3.7)	9.4 (4.0)	0.7 (0.9)
IO 2	9	2.0 (1.6)	8.2 (3.5)	0.2 (0.5)	2.7 (2.3)	7.6 (3.7)	0.5 (0.7)
IO 3	6	0.9 (1.1)	7.1 (4.0)	0.2 (0.5)	1.5 (1.6)	7.7 (4.4)	0.5 (0.8)
IO 4	11	1.6 (1.4)	6.9 (3.1)	0.2 (0.4)	2.1 (1.8)	6.6 (3.4)	0.4 (0.7)
IO 5	11	0.4 (0.7)	3.8 (2.4)	0.1 (0.3)	0.6 (0.9)	4.4 (3.2)	0.3 (0.5)
IO 6	9	0.5 (0.7)	4.7 (2.6)	0.1 (0.3)	0.8 (1.1)	5.6 (3.8)	0.3 (0.6)

Note: ACH= air exchange rate per hour, MV=Mechanical Ventilation

Table 3-10 Costs for the intervention options.

Control Options	Cost[a]		
	Installation	Annual maintenance	Mask & Respirator Annually
IO 1	$2,500	$2,000	$2,135
IO 2	$6,000	$1,175	$2,135
IO 3	$2,500	$2,000	$2,675
IO 4	$3,900	$1,100	$2,135
IO 5	$3,900	$1,100	$2,675
IO 6	$6,000	$1,175	$2,675

[a] According to the study by Ko et al. 2001, and WHO 2009

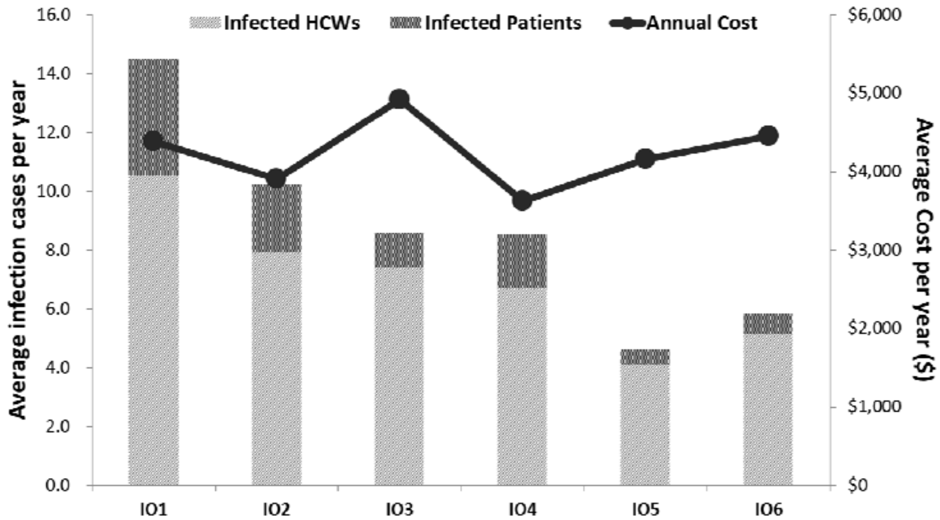


Figure 3-7 Number of infection cases and cost of different intervention options.

3.4.4 Distributive Risk Model

A total of 5,000 iterations were conducted for each segmentation scenario, and the following statistics were calculated and summarized in Table 3-11: mean number of infection cases during the stay of the infectious patient, standard error of the mean, percentage of susceptible patients being infected (infection risk), and *p*-values of the *t*-tests on the means with respect to the one-box model.

The results shows that, the infection risk is largest in the box where the infectious patient sitting, and the susceptible patients sitting closer to the infectious source are clearly under more risks of infection. The four-box model shows that, the infection risks in the adjacent boxes (box 2 and 3) are the second largest. In the box where the source patient is sitting, the infection risk is (0.0213) three times larger than the average of the whole waiting room (0.0055).

According to the results, assuming complete air mixing will underestimate the infection risk. Since the TB intervention regulations are based on the analysis of the whole waiting room without considering the distributive infection risk, it may put patients under more risk of infection. The space segmentation method appropriately models the variation of infection risk with regard to the relative location to the source patient. Also, to achieve higher precision of the distributive infection risk, more number of segmentation is necessary.

This model will be especially useful for resource restricted areas where physical isolation in a separate room is not practical. Instead, virtual isolation policy can be employed that separates patients with regard to a preliminary diagnosis. For example, individuals with TB symptoms can be directed to a corner in the general waiting room, sitting far from others without such symptoms.

Table 3-11 Infection risk using sequential box model with different segmentations.

One-Box Model	**Room Average**					
Mean Infection Cases	0.017					
Standard Error of the Mean	0.0018					
Percentage of Infection	0.0051					
Two-Boxes Model	**Room Average**		**Box 1**	**Box 2**		
Mean Infection Cases	0.0198	(p=0.305)	0.0198	0		
Standard Error of the Mean	0.0020		0.0020	0		
Percentage of Infection	0.0059		0.0119	0		
Four-Boxes Model	**Room Average**		**Box 1**	**Box 2**	**Box 3**	**Box 4**
Mean Infection Cases	0.0182	(p=0.650)	0.0178	0.0002	0.0002	0
Standard Error of the Mean	0.0019		0.0019	0.0002	0.0002	0
Percentage of Infection	0.0055		0.0213	0.0002	0.0002	0

3.4.5 Model Verification

In addition to the preliminary study in section 3.4.2, we performed a sensitivity analysis on the distributions of the infection cases. The assumptions in section 3.4.3 were used in this part. Note that all of the results resemble Poisson distributions with different means, which is consistent with the distribution from a previous study (Ko et al., 2001).

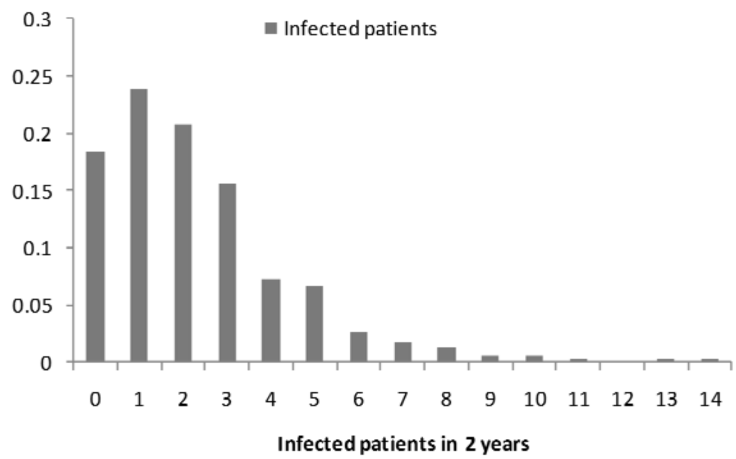


Figure 3-8 Distribution of infected cases among susceptible patients in 2 years

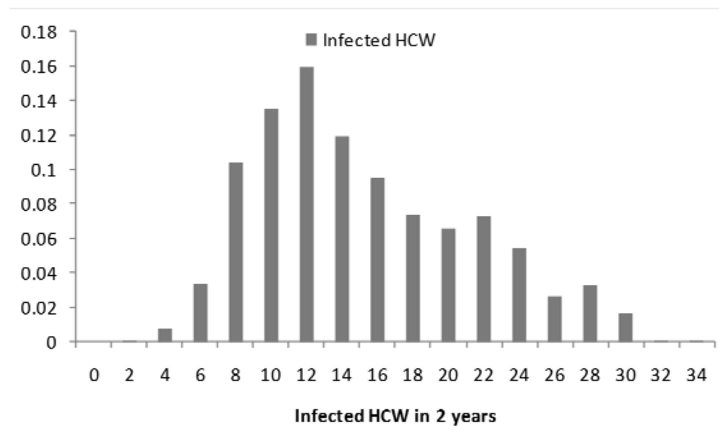


Figure 3-9 Distribution of infected cases among HCWs in 2 years.

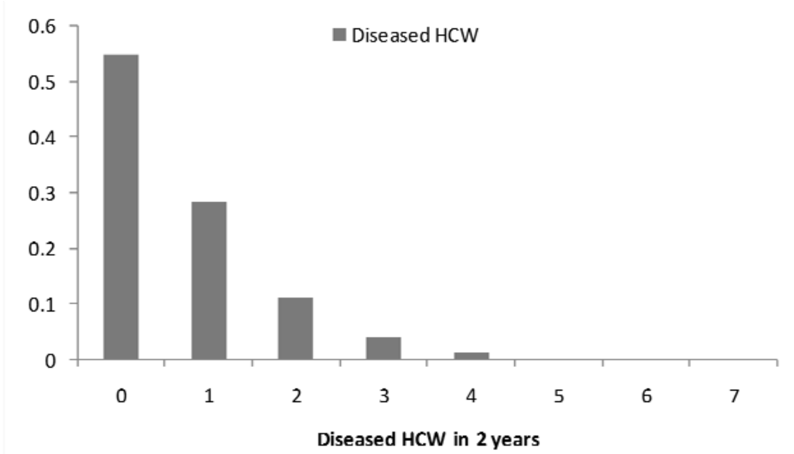


Figure 3-10 Distribution of diseased cases among HCWs in 2 years.

3.5 Application

3.5.1 Frequency of the Screening Program

Screening frequency is extremely important since the existence of a HCW that is positive for TB is an even more dangerous infectious source than a patient. In this section we estimate the impact of screening frequency on effectiveness. The conditions of the health care clinic under study are as follows: a ventilation of 6 ACH is used, all the patients are required to wear a surgical mask, and HCWs wear a particulate respirator. Two screening frequencies are studied: 10 weeks (ongoing transmission setting) and 1 year (medium risk setting). The results are shown in Table 3-12. A reduction in the number of infection/diseased cases was noticed. When a 10 week screening frequency was applied, the number of patient infection cases was 0.05 less in the first year and 0.6 less in the second year, as compared to annual screening. The difference was even larger for HCW infections. When HCWs were screened every 10 weeks instead of annually, the infection and diseased cases were reduced by 0.3 and 0.02 respectively in the first year; 0.8 and 0.05 respectively in the second year.

Table 3-12 Infection cases with different screening frequency for HCWs.

Screening frequency	1st year			2nd year		
	Patient Infected	HCW Infected	HCW Diseased	Patient Infected	HCW Infected	HCW Diseased
1 year	0.86 (1.11)	7.08 (3.99)	0.19 (0.49)	1.46 (1.65)	7.67 (4.39)	0.50 (0.75)
10 weeks	0.81 (0.97)	6.80 (3.14)	0.17 (0.44)	0.87 (1.01)	5.88 (2.88)	0.45 (0.69)

3.5.2 HIV Isolation Strategies

In this section, all patients are required to wear a surgical mask during the whole time they are in the hospital. Both the general waiting room and the isolation rooms ventilate at an equivalent rate of 6ACH with no ventilation method specified. For each intervention option as explained in Table 3-4, Monte Carlo simulation was performed using Matlab with 5,000 iterations, each representing 365 days. The results for the number of infection cases are presented in Table 3-13 and Figure 3-11 for HIV+ and HIV- patients respectively.

The results show that IS2 and IS3 are better options due to fewer infection cases among HIV+ patients. It suggested that at least HIV+ patients should be isolated during the admission into the hospital, and depending on capacity, HIV- patients could either be isolated separately or isolated together with HIV+ patients. Moreover, at different levels of equivalent ACH, assuming only HIV patients were isolated, the corresponding number of infection cases are calculated and illustrated in the Figure 3-12. It again suggests that higher than 12 ACH may not be necessary for the given example.

Table 3-13 Annual infection cases for HIV+ and HIV- patients under different isolation options.

Isolation Options	Room Allocation	Total infection		HIV- infection		HIV+ infection	
IS 1	1 general waiting room	0.37	(0.17)	0.33	(0.16)	0.04	(0.02)
IS 2	1 HIV (HIV+ and HIV-) isolation room	0.30	(0.15)	0.30	(0.15)	0.01	(0.01)
IS 3	1 HIV+, 2 HIV- isolation rooms	0.29	(0.17)	0.28	(0.17)	0.01	(0.01)
IS 4	3 isolation rooms regardless of HIV status	0.21	(0.12)	0.18	(0.11)	0.03	(0.02)

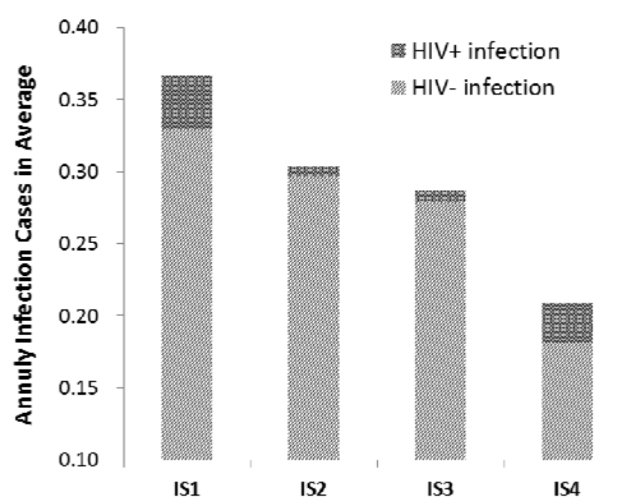


Figure 3-11 Number of infection cases for HIV- and HIV+ patients under different isolation options.

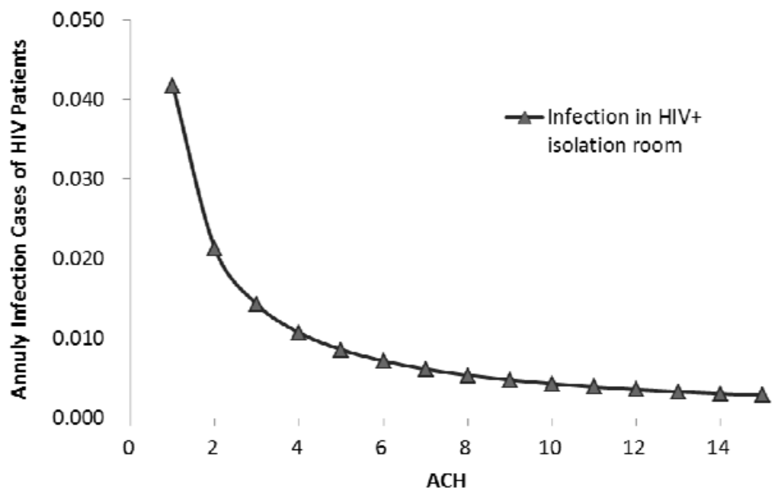


Figure 3-12 Annual number of TB infection cases co-infected with HIV vs. equivalent air exchange per hour in isolation room.

3.6 Conclusion

We developed a model that estimates the infection risk by considering the possible intervention options of: mechanical ventilation, HEPA, UV light, surgical mask, and particulate respirator, etc. Based on this model, the study provided a comprehensive analysis of MTB infection risks under different environmental and managerial control strategies. The study will help hospitals in designing their personalized cost-effective intervention strategy based on their specific situation. This project is in collaboration with the CDC.

A distributive model with space segmentation was also studied to explore whether there is a variation in infection risk with regards to the relative location to the infection source. The model appropriately captures the variation of infection risk with regard to the relative location to the source patient, and shows that susceptible patients sitting closer to the infectious sources are at greater risk of infection. For more accurate modeling of distributive infection risk, the number of segmentation could be increased.

We are currently collaborating with the CDC to gather experimental data from hospitals such that validity of the risk model may be further assessed by practice. Ultimately, the model will assist policymakers in hospitals and clinics to determine the appropriate strategy that match their need.

Chapter 4

Association of Food Environment and Food Retailers with Obesity in U.S. Adults

Abstract

Food environment, including food retailer type, has been shown to be associated with obesity rate. We estimate the marginal effect on obesity rate in U.S. adults based on the addition of a new food retailer type in a geographic region. Nonlinear parametric regression was used on publicly available data (year=2009) at the county level. Obesity rate increased in supercenter and convenience store density and decreased in grocery store and specialty food store density. The marginal measures estimated in this work could be useful in identifying regions where interventions based on food retailer type would be most effective. We also develop an application of the model to assist local governments and foundations to strategically design a food store establishment plan.

4.1 Introduction

4.1.1 Obesity Prevalence and Food Desert

The prevalence of obesity in the United States has increased dramatically over the past two decades (Flegal et al., 2002; Nguyen and El-Serag, 2010). In 2008, 32 states had an obesity prevalence greater than 25%, while in 1990 no state had a prevalence greater than 15% (Nguyen and El-Serag, 2010). In 2010, 35.7% of U.S. adults and 16.9% of children and adolescents were obese (Ogden et al., 2012). Obesity ranks as the second leading "actual cause" of death in U.S., which accounts for 15% of all preventable deaths (McGinnis and Foege, 1993). It substantially increases the risks of heart disease, stroke, type 2 diabetes, and certain types of cancer (NIH, 1998). Individuals who are obese have significantly worsened health-related quality of life, and are also more likely to have a functional limitation (Fontaine and Barofsky, 2001; Swallen et al., 2005).

In addition, a higher incidence of obesity creates an additional burden on society. The medical costs for people who are obese were $1,429 higher than those of normal weight, and nationally, obesity was responsible for almost $40 billion of increased medical spending in 2006 (Finkelstein et al., 2009). Roughly half of such spending is financed by taxpayer dollars in the form of Medicare and Medicaid (Finkelstein et al., 2009).

Individuals undergoing limited accessibility to healthy food retailers are geographically aggregated. "Food deserts" refer to areas *"... in the United States with limited access to affordable and nutritious food, particularly such an area composed of predominantly lower income neighborhoods and communities*" (Pub L No. 110-224, 2008). According to a national sample study, counties classified as food deserts generally have lower education level, higher poverty rates, lower median family income, greater percentage of residents living in rural regions, lower population density, greater median age, and a greater number of small grocery and convenience stores per capita as compared to counties not classified as food deserts (Morton and Blanchard, 2007).

4.1.2 Food Environment in Relation to Obesity

While obesity is caused by a complex interaction of the environment, genetic predisposition, and human behavior (Nguyen and El-Serag, 2010), several studies have found that disparities in the local food environment are a major contributor to obesity (Block and Kouba, 2006; Caspi et al., 2012; Chen et al., 2010b; Evans et al., 2010; Franco et al., 2008; Larson, Story, and Nelson, 2009; Pearce et al., 2007; Walker, Keane, and Burke, 2010). The local food environment is a product of the distribution of food retailers by type, the relative prices of different food products, and car ownership (Morland et al., 2002; Nguyen and El-Serag, 2010). Minority and lower-income populations have greater difficulties in accessing high-quality fresh food, and the influence from food environment tends to be higher than their counterparts (Chen et al., 2010b; Moore and Roux, 2006; Morland et al., 2002; Powell et al., 2007; Walker, Keane, and Burke, 2010). Further, persons living in rural areas have fewer supermarkets and grocery stores, and generally need to travel longer distance to stores, as compared to those live in metropolitan areas (Liese et al., 2007; Menzies et al., 2000; Michimi and Wimberly, 2010; Moore et al., 2008; Morton and Blanchard, 2007; Powell et al., 2007).

Supermarkets and large grocery stores, which offer wide variety of high-quality foods at lower prices, has been extensively studied in relation to obesity prevalence (Block and Kouba, 2006; Krukowski et al., 2010); health outcomes improve with increasing accessibility to supermarkets and grocery stores (Brown et al., 2008; Chen et al., 2010a; Moore et al., 2008; Morland, Diez Roux, and Wing, 2006; Morland and Evenson, 2009). On the other hand, increased access to convenience stores is generally associated with lowered fresh food intake and poorer health conditions (Brown et al., 2008; Galvez et al., 2009; Jago et al., 2007; Morland, Diez Roux, and Wing, 2006; Morland and Evenson, 2009). Studies of other types of stores, such as specialty food stores, are relatively limited, and differ in their conclusions (Block and Kouba, 2006; Evans et al., 2012; Galvez et al., 2009; Moore and Roux, 2006; Morland and Evenson, 2009).

Affordability of food is another important factor since the changes in relative prices of different foods affect demand. The consumption of energy dense and nutrient poor foods by low-income groups may be explained in part by those foods being less expensive and most resistant to inflation than nutrient rich foods (Monsivais and Drewnowski, 2007). The relative difference in cost between healthy food and unhealthy food is higher in lower-income areas (Block and Kouba, 2006; Cummins and Macintyre, 2002). Further, a study in Los Angeles and Sacramento states that cost of a healthier basket of food accounts for a significant portion of low-income consumers' food budgets (Jetter and Cassady, 2006). Some have suggested that taxing energy-dense, low nutrition food while providing subsidies to healthy food alternatives may help to control the increasing obesity rate (Powell, Han, and Chaloupka, 2010).

The cause of the disparities in accessibility and affordability is a rather complex problem. For a variety of reasons, major food retailers do not necessarily establish stores in poor communities (Cummins and Macintyre, 2002) and the trend toward supermarket chain stores moving locations from urban to suburban areas, effectively creates “food deserts” in inner cities (Block and Kouba, 2006). The lack of convenient transportation results in travelling significant distances for both rural low-income and urban residents in order to have access to healthy and affordable food retailers (Clarke, Eyre, and Guy, 2002; Eisenhauer, 2001; Sharkey and Horel, 2008). As a result, people living in those areas have to rely on small grocery or convenience stores that usually supply a limited range of food at premium prices. For this reason, car ownership has become a popular measure for the access, particularly for suburban and rural residents (Handy and Clifton, 2001).

The impact of food assistance programs has also been studied. It is suggested that the The Supplemental Nutrition Assistance Program (SNAP, formerly known as the Food Stamp program or FSP) participation may be positively related to the incidence of obesity. The program was associated with a 9.1% increase in the probability of current obesity among women in low-income group, while no such relationship was found among low-income men (Chen, Yen, and Eastwood, 2005; Gibson, 2003). Townsend et al. (Townsend et al., 2001) argued that SNAP might have caused unhealthy eating habits among the participants which results to the weight gain, and Chen et al (Chen, Yen, and Eastwood, 2005) further argued that the difference among men and women was due to that the program not distinguishing between genders. In light of those studies, food assistance programs should carefully review the program, and think about how to overcome the food-insecurety efficiently while encouraging healthy lifestyles (Just, 2006).

4.1.3 Review of Literature

Many studies have provided empirical evidence of the role of disparities and access to healthy foods. Larson et al. (Larson, Story, and Nelson, 2009) did a systematic review of the literature on disparities in food access from 1985 to 2008, and identified several limitations of current studies including effective policy actions and evaluation strategies to promote the equitable access to healthy foods (Larson, Story, and Nelson, 2009). There are several limitations with current studies, however, including:

1) Geographic level of study

The majority of previous studies on food environment and health outcomes have focused on selected counties or certain regional areas with data collected from a selective population (Brown et al., 2008; Chen et al., 2010a; Galvez et al., 2009; Moore et al., 2008; Morland and Evenson, 2009), while very few have been conducted at a the national level (Michimi and Wimberly, 2010). Local-level studies are useful for their respective regions, but it can be difficult to draw general conclusions that apply across the U.S.

2) Store classifications

Classification of store types varied from study to study in terms of major products, store sizes, number of employees, and ownership. In many studies, there was no clear definition of store type. A large number of studies combined supercenters and supermarkets together in the same class (Chen et al., 2010a; Michimi and Wimberly, 2010; Moore and Roux, 2006), while other studies divided supermarkets into large chain store and private groceries differentiated by name or number of employees (Brown et al., 2008; Moore and Roux, 2006). According to the 2007 version of North American Classification System (NAICS), a standard in classifying business establishments, there is a clear distinction between supermarkets and supercenters. The former is primarily engaged in retailing a general line of foods, while the later also sells merchandise such as apparel, furniture, and appliances. It is possible that the lack of consistency of store types in different studies is one of the reasons for the inconsistent conclusions drawn.

3) Metropolitan versus non-metropolitan areas

Metro and non-metro areas differ in population density, public transportation infrastructure, and socioeconomic distribution. A limited number of studies have compared these two types of areas (Michimi and Wimberly, 2010; Morton and Blanchard, 2007; Powell et al., 2007). According to these studies, non-metro or rural areas underwent a higher risk of becoming a food desert (Michimi and Wimberly, 2010; Morton and Blanchard, 2007), and had a significantly fewer number of chain supermarkets as compared to the urban areas (Powell et al., 2007). In addition, while the obesity rate was positively associated with distance to supermarkets in metro areas, no such relationship was found in non-metro areas (Michimi and Wimberly, 2010). The authors therefore concluded that obesity and food environments in non-metro areas are likely to be determined by a more complex set of socioeconomic factors; no further studies were conducted to test such hypothesis, however.

4) Statistical model

Logistic regression is one of the more popular models adopted in the literature in this area, and estimates the odds ratio as compared to a reference group (Brown et al., 2008; Galvez et al., 2009; Michimi and Wimberly, 2010; Moore et al., 2008; Morland, Diez Roux, and Wing, 2006; Morland and Evenson, 2009). For example, when logistic regression is used to examine the association of supermarket and obesity, the result is interpreted as the odds ratio of obesity between areas with and without supermarkets. Some studies have also considered the interaction between store variable and other factors, such as chronic condition (Brown et al., 2008). Results using a logistic model are straightforward to interpret for group comparisons. However, as limited by the nature of discrete variables in the model, the method is not very useful for applications such as estimating the marginal effects when one additional store type is added. Linear regression model has also been used for some studies. Chen et al. (Chen et al., 2010a) used a conventional linear model with an added spatial multiplier. However, the adjusted R^2 value was only 0.06. One major deficiency of linear models is the default assumption of a constant effect across the whole range of explanatory variable.

4.1.4 Objective

In this study, the obesity rate of U.S. adults is modeled using county level data in a nonlinear regression in order to identify nationally representative insights on food environmental effects. Four types of food retailers are included: supercenters, supermarkets, convenience stores and specialty food stores. The relative price of healthy and unhealthy food, proximity and car ownership, food assistance, as well as the socioeconomic variables are also included into the model. Metro and non-metro counties are analyzed separately and compared accordingly. The marginal effect of having additional food retailer types on obesity is estimated from the nonlinear regression model. Furthermore, a case study is developed in order to show how to apply the results of the model to solve for an appropriate intervention policy in a specified local area.

4.2 Methodology

4.2.1 Data Description

County-level data from the U.S. Department of Agriculture (USDA), American Community Survey (ACS), Centers for Disease Control and Prevention (CDC), and the U.S. Census Bureau County Business Patterns (CBP) for food environment, food assistance programs, and demographic characteristics by county type (metropolitan or non-metropolitan) was used. The response variable for the study is obesity rate in 2009; the variable list and corresponding summary statistics are shown in

Table 4-1. Moreover, in order to validate the final model, obesity rate in 2008 was used. The corresponding summary statistics for the validation datasets are shown in Table 4-2.

Table 4-1 Obesity modelling data and descriptive statistics (2009).

Variable	Description	Metro (N=1069) Mean	(SD)	Non-metro (N=1972) Mean	(SD)	Data Source
Response Variable						
Obesity Rate	Adult obesity rate	29.51	(4.21)	30.57	(3.92)	CDC, 2009
Food environment						
Low Access to Store	Households, no car & low access to supermarket or large grocery store (%)	2.38	(1.57)	3.48	(3.72)	USDA, 2010
Supercenters	Supercenters & club stores/1,000 pop	0.02	(0.01)	0.02	(0.02)	CBP, 2009
Grocery Stores	Supermarkets and Grocery stores/1,000 pop	0.18	(0.09)	0.33	(0.28)	CBP, 2009
Convenience Stores	Convenience stores/1,000 pop	0.47	(0.19)	0.68	(0.35)	CBP, 2009
Specialty Food Stores	Specialty food stores/1,000 pop	0.06	(0.05)	0.06	(0.09)	CBP, 2009
Farmers' Markets	Farmers' markets/1,000 pop	0.02	(0.03)	0.04	(0.08)	USDA, 2009
Price Ratio (Milk to Sodas)	Price of low-fat milk/price of sodas	0.92	(0.14)	0.90	(0.12)	USDA, 2010
Food Assistance						
SNAP Authorized Stores	SNAP-authorized stores/1,000 pop	0.60	(0.26)	0.90	(0.49)	USDA, 2008
WIC Authorized Stores	WIC-authorized stores/1,000 pop	0.16	(0.09)	0.34	(0.30)	USDA, 2008
Demographic Characteristic						
Race Ratio	White people per non-white people	7.75	(9.50)	13.13	(13.64)	US Census, 2010
Population Density	Population Density (per square mile of land area)	671.19	(2909.66)	45.50	(102.39)	US Census, 2010
Median Age	Median age	37.46	(4.07)	40.64	(5.10)	ACS-5yr, 2009
Education Less than High School	% Education Less than high school graduate (aged 25-64)	12.31	(5.91)	14.60	(7.69)	ACS-5yr, 2009
Unemployment Rate	Unemployment rate (>15 civilian)	6.93	(2.12)	6.80	(3.52)	ACS-5yr, 2009
Poverty Rate	Poverty rate	14.60	(5.27)	17.62	(6.18)	US Census, 2010
Sex Ratio	males per females of total population	97.98	(7.27)	100.78	(13.40)	ACS-5yr, 2009

ACS: American Community Survey
CBP: U.S. Census Bureau, County Business Patterns
USDA: U.S. Department of Agriculture
CDC: CDC's Behavioral Risk Factor Surveillance System (BRFSS)
* Geographic Level: 26 markets and 9 non-metro census divisions based on Nelson

Table 4-2 Obesity model validation data and descriptive Statistics (2008).

Variable	Description	Metro (N=1069) Mean	(SD)	Non-metro (N=1970) Mean	(SD)	Data Source
Obesity Rate	Obesity rate of persons age > 20	28.86	(3.70)	29.03	(3.74)	CDC, 2008
Supercenters	Supercenters & club stores/1,000 pop	0.02	(0.02)	0.02	(0.02)	CBP, 2008
Grocery Stores	Supermarkets and Grocery stores/1,000 pop	0.27	(0.21)	0.29	(0.26)	CBP, 2008
Convenience Stores	Convenience stores/1,000 pop	0.60	(0.31)	0.63	(0.34)	CBP, 2008
Specialty Food Stores	Specialty food stores/1,000 pop	0.06	(0.07)	0.06	(0.07)	CBP, 2008
Poverty Rate	Poverty rate	15.32	(6.05)	15.42	(6.07)	U.S. Census, Small Area Income and Poverty Estimates 2008

CDC: CDC’s Behavioral Risk Factor Surveillance System (BRFSS)
CBP: U.S. Census Bureau, County Business Patterns

Comparing Table 4-1 and Table 4-2 side by side, we observed that the obesity rate has increased by approximately 1% from 2008 to 2009 for both metropolitan and non-metropolitan counties. The disparity between metro and non-metro counties' obesity rate has increased from 0.2% in 2008 (28.86% for metro and 29.03% for non-metro) to 1% in 2009 (29.51% for metro and 30.57% for non-metro). The densities of grocery and convenient stores have significantly dropped in metro counties from 2008 to 2009, but slightly increased in non-metro counties.

Metro and non-metro information are obtained from the USDA's Economic Research Service — Rural Classification in 2000, as difined by the Office of Management and Budget (OMB). Metro areas are defined for all urbanized areas regardless of total area population. Outlying counties are also classified as metro if they are economically tied to the central counties, as measured by the share of workers commuting on a daily basis to the central counties. Non-metro counties are outside the boundaries of metro areas and have no cities with 50,000 residents or more.

The North American Classification System (NAICS) is used by Census Bureau County Business Patterns (CBP) to classify stores into four types:

1) Warehouse and Supercenters (NAICS: 452910) comprises of establishments known as warehouse clubs, superstores or supercenters primarily engaged in retailing a general line of groceries in combination with general lines of new merchandise, such as apparel, furniture, and appliances. Major companies include Sam's Club (Walmart); Costco Wholesale; BJ’s Wholesale Club; and Meijer.

2) Supermarkets and Grocery stores (NAICS: 445110) generally known as supermarkets and smaller grocery stores primarily engaged in retailing a general line of food, such as canned and frozen foods; fresh fruits and vegetables; and fresh and prepared meats, fish, and poultry. Major companies include Kroger, Safeway, and SUPERVALU (all based in the U.S.).

3) Convenience stores (NAICS: 445120 and 447110) are food marts (except those with fuel pumps) primarily engaged in retailing a limited line of goods that generally includes milk, bread, soda, and snacks. Such stores may or may not also engage in retailing automotive fuels. Major companies include 7-Eleven and Sheetz.

4) Specialty food stores (NAICS: 4452xx) are primarily engaged in retailing specialized lines of food, such as retail baked goods, meat and seafood markets, dairy stores, fruit and vegetable markets, confectionery and nut stores, and others. Examples include butcher shops, bakery stores, coffee and tea stores, etc. No major companies dominate the industry.

Two food assistances programs are considered, the Supplemental Nutrition Assistance Program (SNAP) and the Special Supplemental Nutrition Program for Women, Infants, and Children (WIC) Program. WIC does not include direct distribution contractors in Mississippi (MS), while Vermont (VT) uses home delivery contractors for most foods, and began authorizing WIC retail stores in 2009 for fruits and vegetables. Because of these differences, data from these two states were excluded.

Counties with population densities in the range of 0.1 to 2,500 persons per square mile were used. In addition to MS and VT being excluded due to differences in SNAP policy, 30 counties were removed for missing values, and 3 counties that were identified as outliers were excluded, leaving 1,015 observations for metro areas and 1,930 for non-metro areas.

4.2.2 Model

In this section, the modeling process will be explained step by step, where step 0 is the preparation or transformation of the response variable and step 1-3 are the modeling procedures.

Step 0.Transformation on Response Variable

Since the rate of obesity is slightly negatively skewed (Skewness is -0.35 and -0.28 for non-metro and metro areas respectively), a Box-Cox transformation (Box and Cox, 1964) was used. It is expressed by,

$$y' = \begin{cases} \frac{y^{\lambda}-1}{\lambda} if \lambda > 0 \\ \log(y)\ if \lambda = 0 \end{cases}.$$

The value of λ was chosen that maximized the log-likelihood function,

$$\ln L(\lambda) = -\left(\frac{N}{2}\right) ln2\pi - \left(\frac{N}{2}\right) ln\sigma_{\lambda}^{2} - \sum_{i=1}^{N} \frac{\left(y_i^{(\lambda)} - \mu_{\lambda}\right)^2}{2\sigma_{\lambda}^{2}} + (\lambda - 1) \sum_{i=1}^{N} ln y_i,$$

where N is the number of observations and i=1,2,….N.

Step 1. Step-wise Regression

Table 4-1 contains the feasible explanatory variables. A stepwise multiple-linear regression analysis is applied to select the best independent variables to predict obesity rate (transformed) for metro and non-metro counties. The significance level for entry (SLE) and the significance level for staying in the model (SLS) are 0.15 and 0.05 respectively in the stepwise selection. For each of the identified significant variables, the residual plot was examined to determine if a curvature problem exists. Meanwhile, variance inflation factor (VIF) is used to check for multicollinearity, and a variable with VIF values greater than 10 may require further investigation.

Step 2. Spline Regression

Based on the results from step 1, the significant explanory variables are included for a spline regression, which transforms each significant explanatory variable so that an additive model can be applied. According to the shape of the residual plots, two types of transformations were applied. For variables with residuals showing a curventure relationship to the obesity rate, a spline transformation without monotonic constraints was applied; otherwise, a monotone spline regression was applied.

A B-spline transformation (De Boor, 1978; van Rijckeveorsel, 1982) of each independent variable is optimaly derived using SAS. To begain with, a B-spline basis of the specified degree (n) with the specified interior knots (k) (De Boor, 1978) was created. The basis has n+k+1 columns (1 for intercept). The optimal spline is a linear combination of B-spline basis columns. By performing such transfomrations on each of the independent variables, the predicted-values for the dependent variable is simply the sum of the transformed independent variables. Ordinary least-squares regression is used for the determination of the linear coefficients. B-splines are equavalent to constructing basis for piecewise polynomials, but is more computationally accurate and efficient. In cases where a monotone relatinship is required between the independent and dependent variables, monotonicity constraints are placed on the coefficients of the B-spline basis (Winsberg and Ramsay, 1980).

Step 3. Nonlinear Parametric Model

A nonlinear parametric model was applied to approximate the shape of the spline transformation from step 2. The objective is to achieve a model with good fitting accuracy using a fewer number of modeling coefficients as compared to the Spline model. Four common functions were selected based on the shape of the spline curve to model each of the independent variables: i) linear, ii) logarithmic, iii) quadratic, and iv) sigmoid (s-shaped) or its mirror. Judging from the later results, those four functions are sufficient for this modeling problem. There are many sigmoidal functions to capture the eglongted S-shaped model including logistic, Gompertz, Richards, Schnute, and Stannard curves. The logistic growth model was chosen to model the S-shaped function due to the relatively low number of parameters required for estimation. The model is expressed as:

$$y = g_0 + g_1 \cdot \frac{1}{1+e^{-(x-\lambda)\cdot\alpha}} + u,$$

where y is the transformed value of x plus a standard uniformly distributed residual term, u. Four parameters need to be estimated: the rate of change α, the value where the rate of change reaches its maximum λ (i.e., the inflection point), and the lower and upper asympptotes g_0 and $g_0 + g_1$. It is assumed that all y values follow a normal distribution.

The specified models are fit by maximizing an approximation to the likelihood integrated over the random effects using SAS. An Adaptive Gaussian Quadrature method is used to approximate the likelihood (Pinheiro and Bates, 1995). The parameters are estimated iteratively to achieve the opimal value. Quasi-Newton methods generally perform well for each predictor on the aspect of fitting the parametric model to a B-splines with monotonicity constraints (Winsberg and Ramsay, 1980). The one exeception is the gender ratio in the non-metro model. In this case, the conjugate gradient methods is applied instead. With the fitted transformation on each independent variable, a multiple linear regression is conducted to predict the obesity rate at each county.

There are several limitations to this study that should be mentioned. First, data on store proximity by distance was not available and hence store density, low accessibility, and percentage of car ownership were used as proxies. Second, it was assumed that individuals in a geographic area shopped in stores in their county and did not cross regions. Previous literature is mixed on the severity of this limitation. Brown et al. (Brown et al., 2008) found no major differences in their results when they included surrounding areas as shopping possibilities as compared to when crossing regions were not allowed. While Chen et al. (Chen et al., 2010b) found that for census tract data the results did matter. Third, restaurant data was not included in the model, which is an important component of the food environment. Finally, publicly available data may be subject to error (Sharkey and Horel, 2008).

4.3 Results

4.3.2 Fitting Results

Step 0.Transformation on Response Variable

A convenient value of λ within the 95% CI range is chosen for the Box-Cox transformation on obesity rate (Figure 4-1). The selected value of λ is 2 for non-metro area, and 1.5 for metro area.

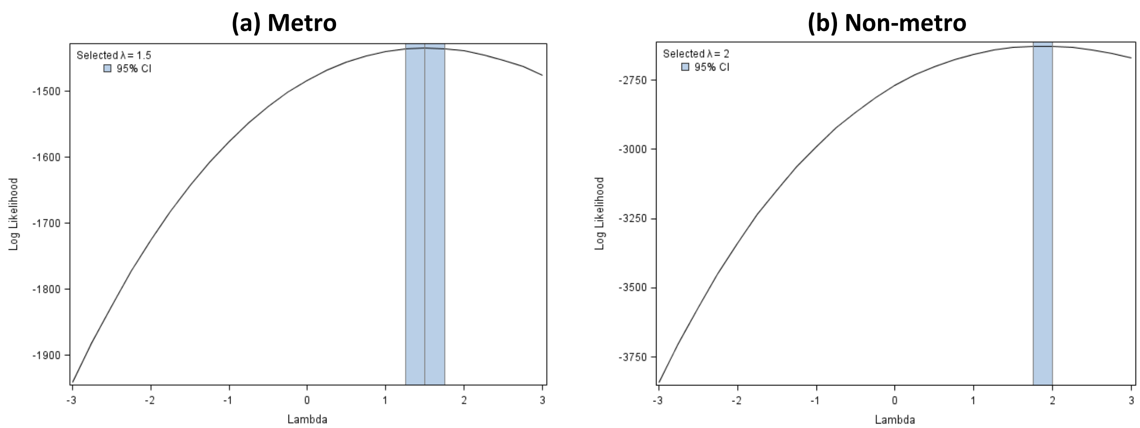


Figure 4-1 Box-cox analysis for obesity rate: log Likelihood versus λ with the 95% confident interval.

Step 1. Step-wise regression

Several factors are found to be associated with obesity rate from the step-wise regression analysis (Table 4-3). The explained variation in transformed obesity rate was 0.395 and 0.312 for metro and non-metro areas respectively. Low access to store (defined as low accessibility and no car ownership) is the major explanatory variable (p<.0001) for obesity rate in metro areas. For non-metro areas, poverty rate is the main explanatory variable followed by the access rate. Also according to the results of VIF, none of the predictor variables listed in Table 4-3 has a value greater than 10. Therefore, the predictors selected from this step are not correlated.

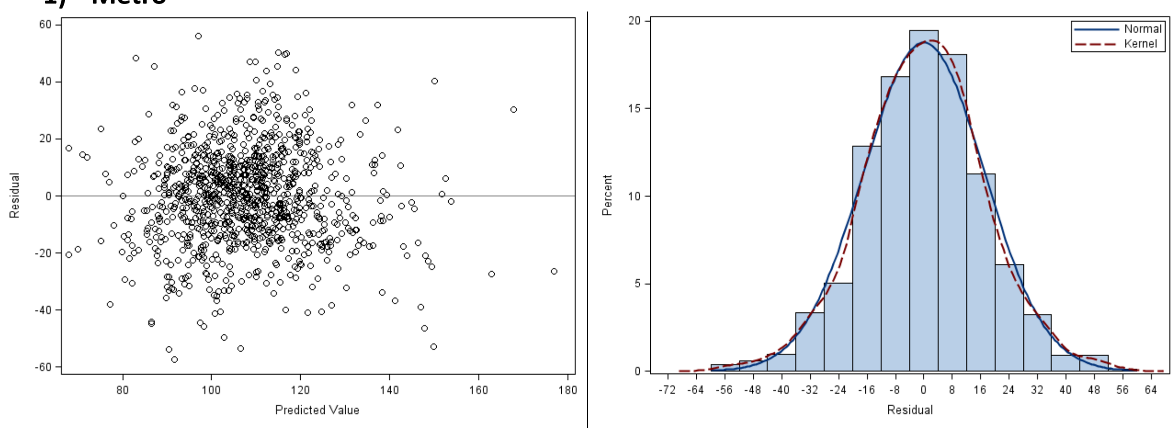
The residuals for obesity rate followed a normal distribution (Figure 4-2). Figure 4-3 showed the plot of residuals for each independent variable for non-metro and metro counties respectively. The plot of residuals versus "median age" revealed a curvature pattern for both areas, and a similar problem is found for population density and price ratio in non-metro areas. In addition, the residuals have a larger deviation at the lower ends of "supercenter", "race ratio" and "population density" for both areas; and "specialty food stores" for non-metro areas. Those results suggested that a linear model might not be appropriate to efficiently capture the variation of obesity rate. Also note that several influential and outlying observations exist.

Table 4-3 Stepwise regression of Box-Cox transformed obesity rate.

Metro Counties (N=1018) df=10					**R^2=0.395**
Significant Variable	**Coefficient**	**Standard Error**	**p-value**	**Partial R^2**	**VIF**
Intercept	87.52	5.92	<.0001		
Low Access to Store	3.02	0.44	<.0001	20.4%	1.58
SNAP Authorized Stores	16.39	2.89	<.0001	7.2%	1.84
Specialty Food Stores	-69.63	11.91	<.0001	4.3%	1.12
Unemployment Rate	2.15	0.30	<.0001	2.3%	1.33
Population Density	-0.01	0.00	<.0001	2.6%	1.26
Convenience Stores	15.82	3.55	<.0001	0.9%	1.53
Grocery Stores	-18.33	7.51	0.017	0.6%	1.30
Race Ratio	0.23	0.06	<.0001	0.5%	1.32
Supercenters	110.53	40.59	0.004	0.5%	1.13
Median Age	-0.36	0.15	0.009	0.3%	1.31
Non-metro Counties (N=1944) df=11					**R^2=0.312**
Significant Variable	**Coefficient**	**Standard Error**	**p-value**	**Partial R^2**	**VIF**
Intercept	347.33	35.62	<.0001		
Poverty Rate	2.54	0.55	<.0001	19.6%	2.38
Low Access to Store	6.47	1.18	<.0001	3.0%	1.55
Price Ratio (Milk to Sodas)	229.53	21.74	<.0001	2.4%	1.35
Unemployment Rate	6.92	0.87	<.0001	2.1%	1.83
Race Ratio	1.03	0.19	<.0001	1.0%	1.31
Median Age	-3.13	0.52	<.0001	1.1%	1.37
Sex Ratio	-0.87	0.18	<.0001	0.8%	1.07
WIC Authorized Stores	34.89	9.23	0.0002	0.5%	1.21
Supercenters	301.34	95.38	0.0016	0.3%	1.12
Population Density	-0.07	0.03	0.0146	0.2%	1.15
Specialty Food Stores	-55.99	25.79	0.0301	0.2%	1.05

* Partial R^2 is the remaining variation explained by the variable given that variables in the previous rows are already in the model.

1) Metro



2) Non-metro

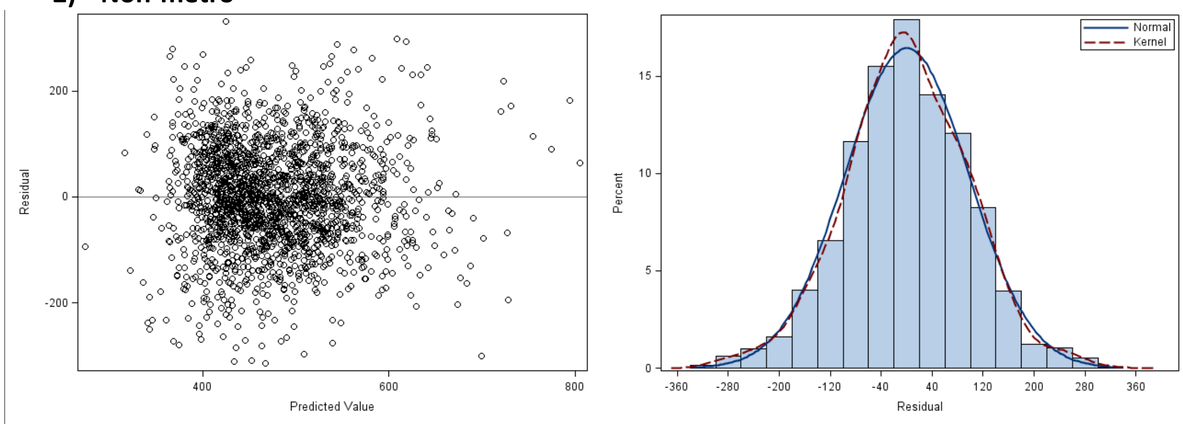
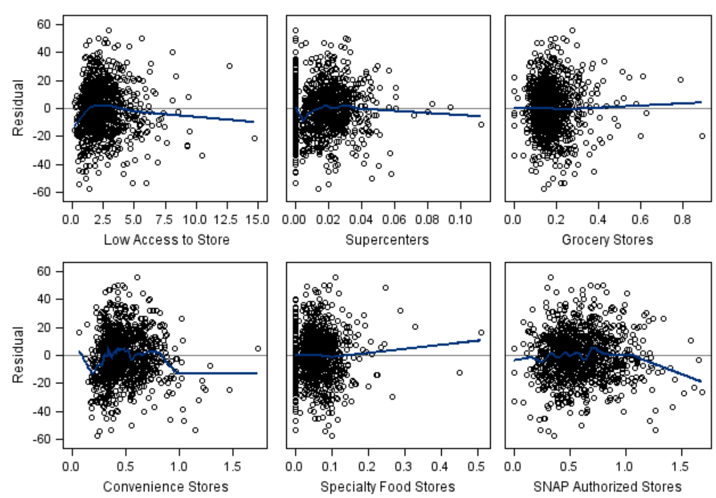
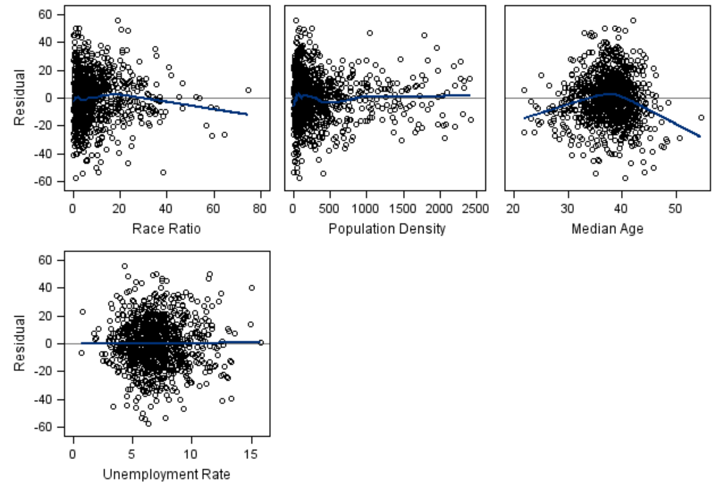


Figure 4-2 Residual plots of transformed obesity rate and the error distributions from step-wise regression.

1) Metro





2) Non-metro

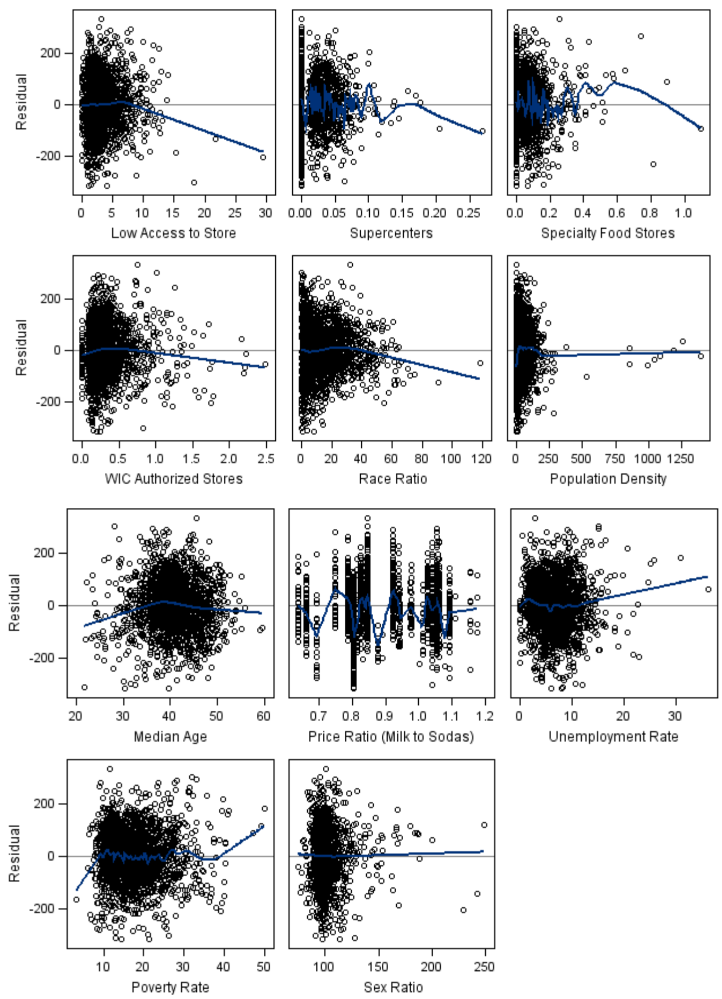


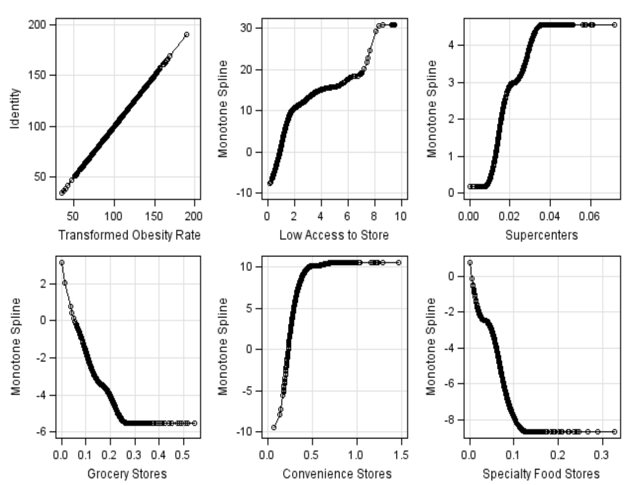
Figure 4-3 Predictors' residual plots from step-wise regression with LOESS smooth curve.

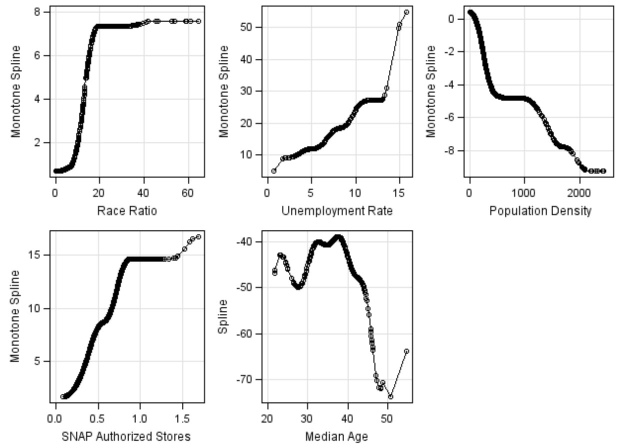
Step 2. Spline Regression

Observations identified as outliers were removed from the data, resulting in 1,984 observations for non-metro areas and 1,000 observations for metro areas. The spline transformation curves for each independent variable are shown in Figure 4-4 with related coefficients summarized in Table 4-4. The transformation column decribes whether or not a monotone constraine was used. Since the degrees of freedom are not always 1 for each independent variable, *t* tests are not uniformly appropriate. Therefore, type II sum of squares and *p*-values based on transformation degrees of freedom (i.e., conservative *p*-values) were reported. The R^2 values are 0.489 and 0.383 for metro and non-metro counties respectively, which improved by approximately 0.1 for both areas as compare to the first step. The residual versus response and predictor plots (Figure 4-5 and Figure 4-6) show that the cuventure patterns in the linear model are much improved with no obvious patterns observed.

We observed that the change of the number and location of knots and the degree of the spline can significantly change the regression results and the smoothness of the curve. Although increasing the number of knots gives the spline more flexibility and hence more closely follows the data, the resulting curve tends to be less smooth. After multiple trials, the accuracy and curve smothness were balanced, and nine knots were evenly placed within the value range and a cubic spline (degree equal to three) was selected.

1) Metro





2) Non-metro

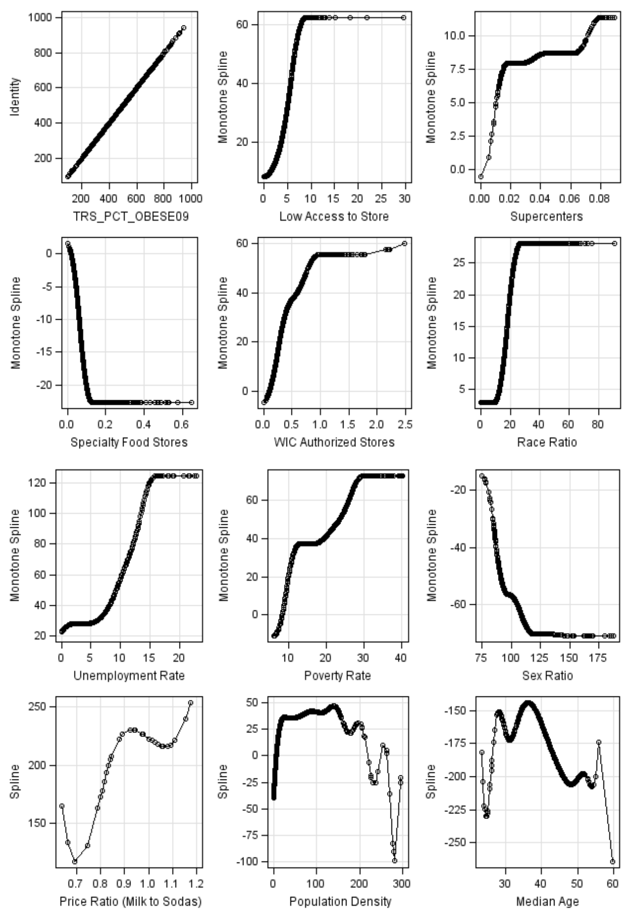
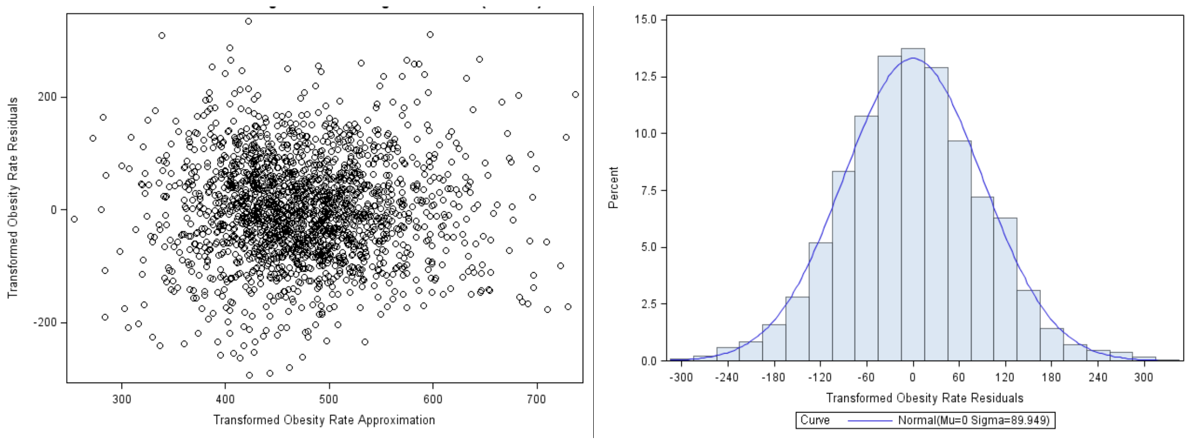


Figure 4-4 Spline transformation for each independent variable.

1) Metro



2) Nno-metro

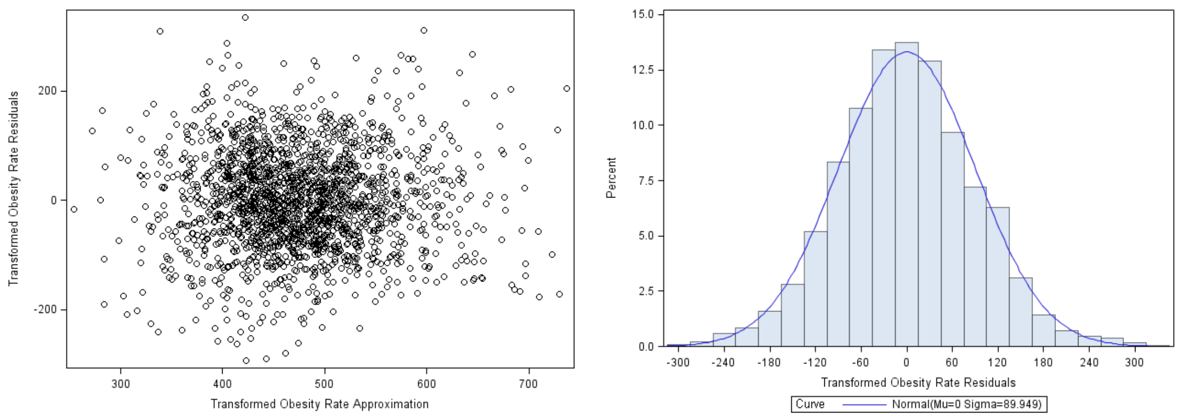
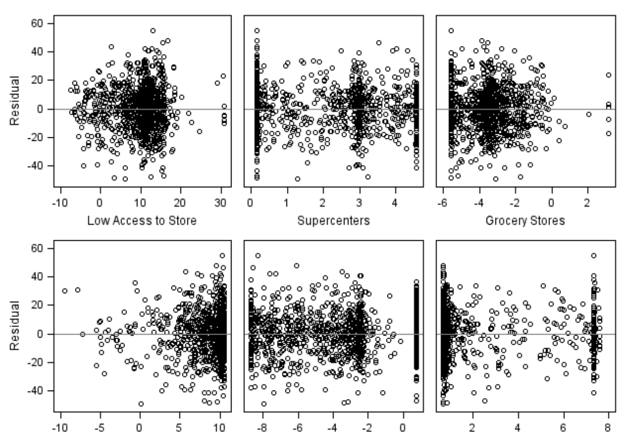
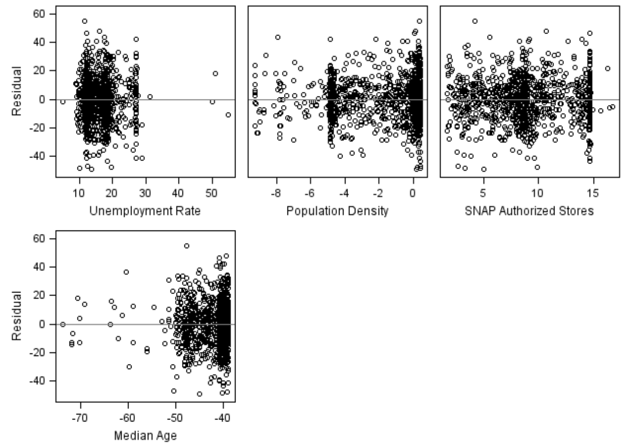


Figure 4-5 Residual plots of transformed obesity rate, and the error distributions from spline regression.

1) Metro





2) Non-metro

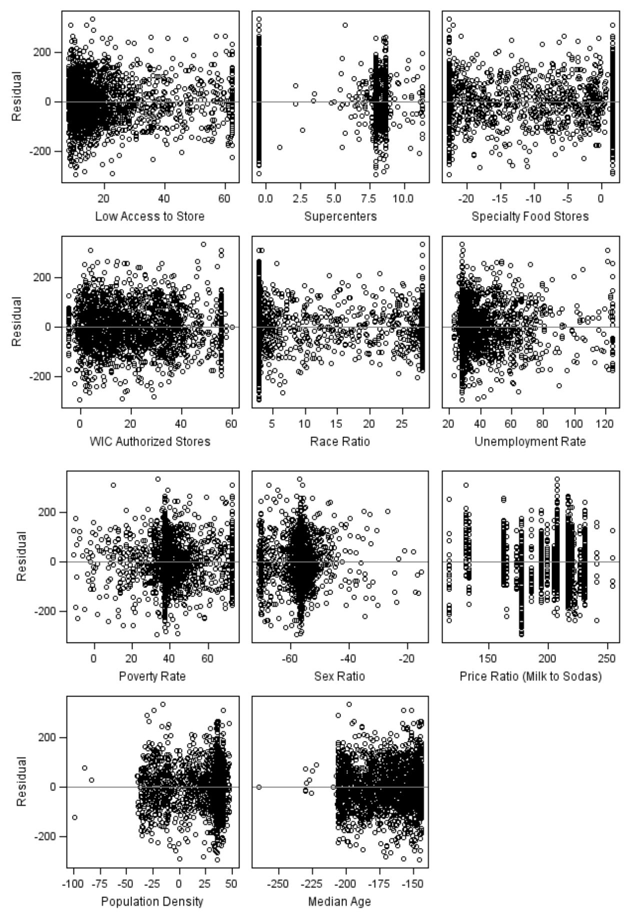


Figure 4-6 Predictors' residual plots from spline regression.

Table 4-4 Spline regression of transformed obesity rate.

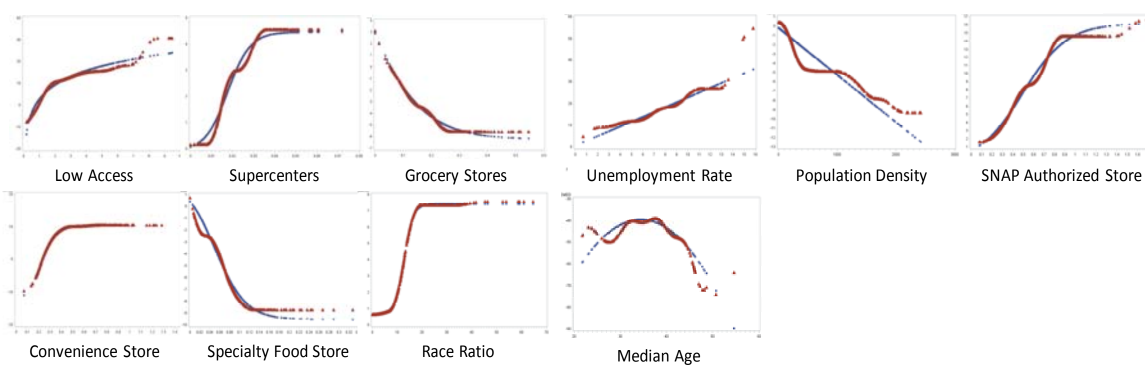
Metro Counties (N=1000)					**R^2=0.489**
Variable	**DF**	**Type II Sum of Squares**	**Mean Square**	**Conservative p**	**Transformation**
Intercept	1	106695	106695	<= .0001	112.49
Low Access to Store	11	31149	2832	<= .0001	Monotone Spline k=9, n=3
Supercenters	11	11472	1043	<= .0001	Monotone Spline k=9, n=3
Grocery Stores	11	9151	832	<= 0.0004	Monotone Spline k=9, n=3
Convenience Stores	11	15976	1452	<= .0001	Monotone Spline k=9, n=3
Specialty Food Stores	11	8575	780	<= 0.0010	Monotone Spline k=9, n=3
Race Ratio	11	13392	1217	<= .0001	Monotone Spline k=9, n=3
Unemployment Rate	11	25386	2308	<= .0001	Monotone Spline k=9, n=3
Population Density	11	15587	1417	<= .0001	Monotone Spline k=9, n=3
SNAP Authorized Stores	11	12294	1118	<= .0001	Monotone Spline k=9, n=3
Median Age	12	31768	2647	<= .0001	Spline k=9, n=3
Model	111	231587	2086.369	<= .0001	

Non-metro Counties (N=1896))					**R^2=0.383**
Variable	**DF**	**Type II Sum of Squares**	**Mean Square**	**Conservative p**	**Transformation**
Intercept	1	748070	748070	<= .0001	353.027
Low Access to Store	11	408562	37142	<= .0001	Monotone Spline k=9, n=3
Supercenters	11	202838	18440	<= 0.0153	Monotone Spline k=9, n=3
Specialty Food Stores	11	214946	19541	<= 0.0097	Monotone Spline k=9, n=3
WIC Authorized Stores	11	513968	46724	<= .0001	Monotone Spline k=9, n=3
Race Ratio	11	352395	32036	<= .0001	Monotone Spline k=9, n=3
Unemployment Rate	11	675052	61368	<= .0001	Monotone Spline k=9, n=3
Poverty Rate	11	391043	35549	<= .0001	Monotone Spline k=9, n=3
Sex Ratio	11	182277	16571	<= 0.0326	Monotone Spline k=9, n=3
Price Ratio (Milk to Sodas)	4	1389220	347305	<= .0001	Spline k=1, n=3
Population Density	12	626515	52210	<= .0001	Spline k=9, n=3
Median Age	12	750973	62581	<= .0001	Spline k=9, n=3
Model	116	9688669	83523.01	<= .0001	

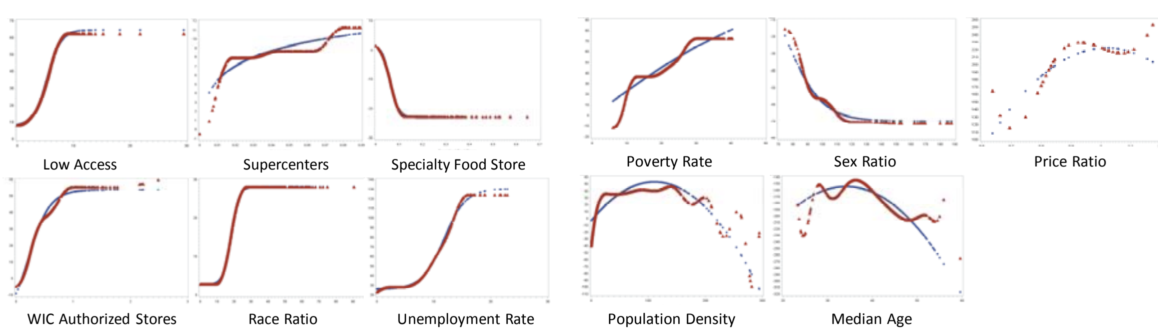
Step 3. Nonlinear Parametric Model

The results of the nonlinear parametric regression are provided in Table 4-5, where all variables included in the final model and its transformation coefficients are listed. The transformation curves are plotted as Figure 4-7, with the spline tranformation in red and nonlinear parametric transformation in blue. The transformation type and parameters are shown for each independent variable that remained significant. In addition, the regression coefficients and resulting *p*-values are provided. The adjusted R^2 values 0.461 for metro counties and 0.342 for non-metro counties. The distribution histograms of the residuals are shown in Figure 4-8. The residual versus predictor plots were also examined (Figure 4-9), and no obvious patterns are observed.

1) Metro



2) Non-metro



Note: Transformation from spline (blue) with parametric fitting curves (red).

Figure 4-7 Fitting curves to the spline transformation for each independent variable.

Table 4-5 Nonlinear parametric regression coefficients.

Metro (N=1000)			***R*²=0.461**
Variable	**Transformation**	**Coefficient**	**p-value**
Intercept	/	112.268	<.0001
Low Access to Store	Log (0,2.885,9.436)	1.018	<.0001
Supercenters	Sigmoid (-0.085,4.558,0.018,186.83)	0.684	0.0446
Grocery Stores	Sigmoid (18.911,-25.128,-0.059,9.56)	0.969	0.021
Convenience Stores	Sigmoid (-12.957,23.373,0.217,15.667)	1.074	<.0001
Specialty Food Stores	Sigmoid (2.357,-11.885,0.051,32.372)	1.072	<.0001
Race Ratio	Sigmoid (0.665,6.765,12.736)	1.142	<.0001
Unemployment Rate	Linear (0.653,2.223)	0.990	<.0001
Population Density	Linear (-0.117,-0.0005)	0.885	0.0006
SNAP Authorized Stores	Sigmoid (-0.916,17.103,0.495,4.741)	1.008	<.0001
Median Age	Polynomial (-1.8591,8.5118,-0.124)	1.004	<.0001
Non-metro (N=1897)			***R*²=0.342**
Variable	**Transformation***	**Coefficient**	**p-value**
Intercept	/	363.426	<.0001
Low Access to Store	Sigmoid (8.021,56.824,5.330,0.817)	1.100	<.0001
Supercenters	Log (0.001,16.658,2.481)	1.270	0.0282
Specialty Food Stores	Sigmoid (2.445,-25.491,0.062,53.909)	0.891	0.0001
WICS Authorized Stores	Sigmoid (-28.121,81.991,0.222,5.386)	0.901	<.0001
Race Ratio	Sigmoid (2.732,25.523,18.138,0.435)	1.025	<.0001
Unemployment Rate	Sigmoid (26.068,104.59,11.761,0.519)	0.978	<.0001
Poverty Rate	Sigmoid (-196.72,331.85,-11.299,0.032)	1.050	<.0001
Sex Ratio	Sigmoid (-26.95,96.64,0.08,75.51)	1.165	0.0024
Price Ratio (Milk to Sodas)	Polynomial (-586.1,1568.75,-778.37)	1.054	<.0001
Population Density	Polynomial (-2.843,1.018,-0.005)	0.853	<.0001
Median Age	Polynomial (-444.34,17.036,-0.500)	1.067	<.0001

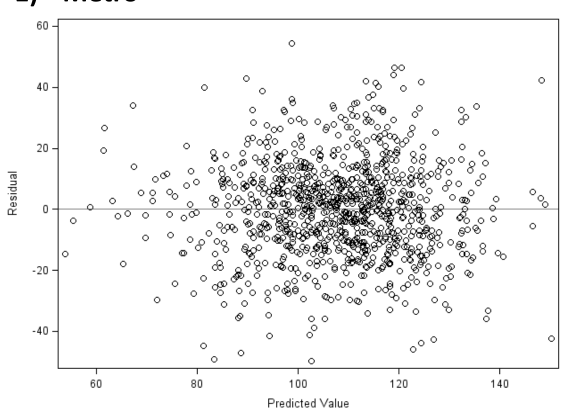
Note: $Linear(a,b) = a + b \cdot x$

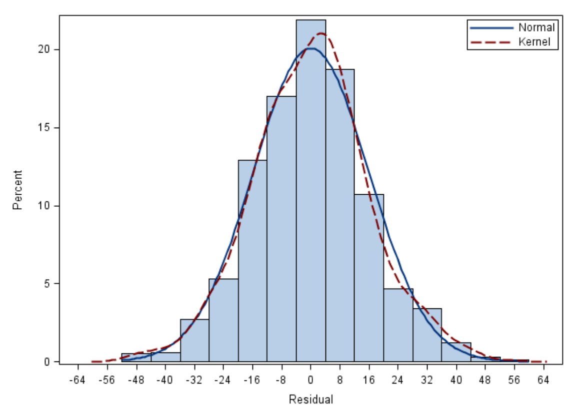
$Polynomial(a,b,c,d) = a + b \cdot x + c \cdot x^2$

$Sigmoid(g_0, g_1, \lambda, \alpha) = g_0 + g_1 \cdot \frac{1}{1+e^{-(x-\lambda)\cdot\alpha}}$

$Log(s,a,b) = a + b \cdot \log(x+s)$

1) Metro





2) Non-metro

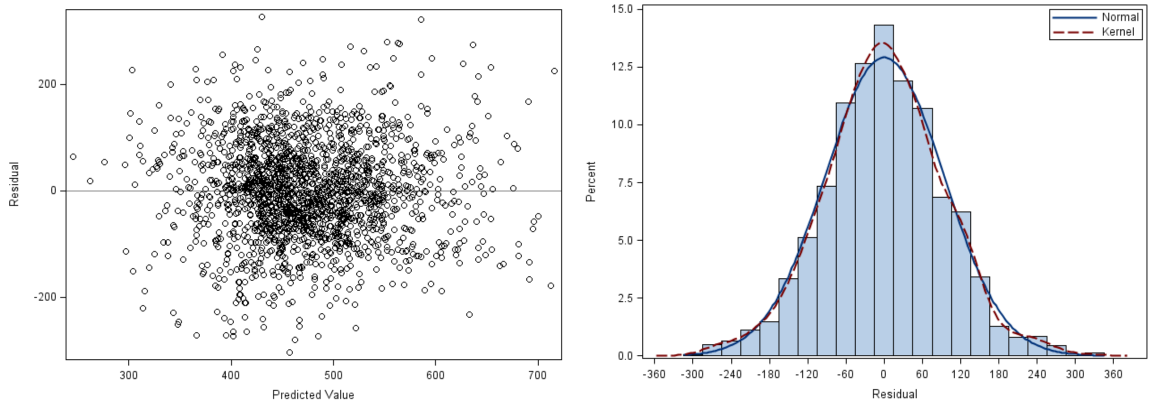
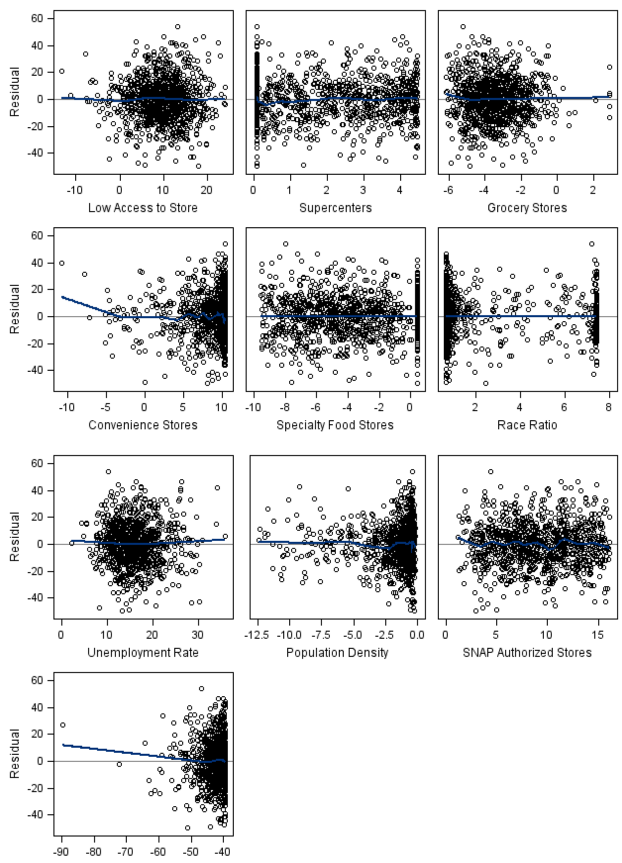


Figure 4-8 Residual plots of transformed obesity rate and the error distributions from nonlinear parametric regression.

1) Metro



2) Non-metro

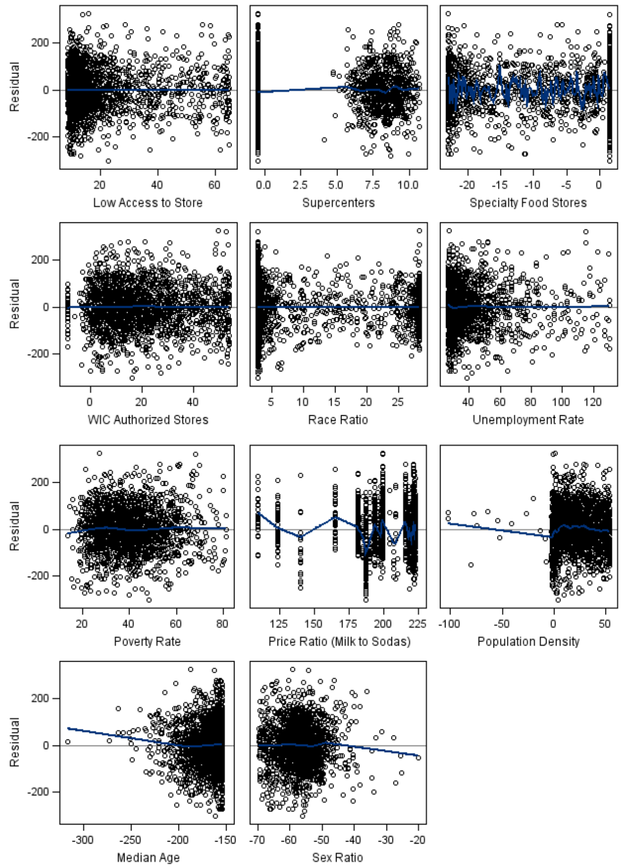


Figure 4-9 Predictors' residual plots from nonlinear parametric regression.

4.3.3 Model Validation

The fitted nonlinear parametric model was validated with an external dataset. Data from 2008 (Table 4-2) were used to validate the model developed by data of 2009 (Table 4-1). More specifically, the external data include 2008 predictors of adult obesity rate, food stores (super centers, grocery store, convenient stores, and specialty food stores) per thousand populations, and poverty rate.

The mean absolute prediction error (MAE) for 2008 is 2.330% for non-metro and 2.284% for metro counties. This accuracy is comparable to that from the training data of 2009, which has prediction MAEs of 2.407% for non-metro and 2.294% for metro areas. The predicted vs. observed county obesity rate are shown in Figure 4-10, for training data and validation data respectively. Once again, the figures show similar prediction accuracy for both datasets and suggest that the fitted model is appropriate.

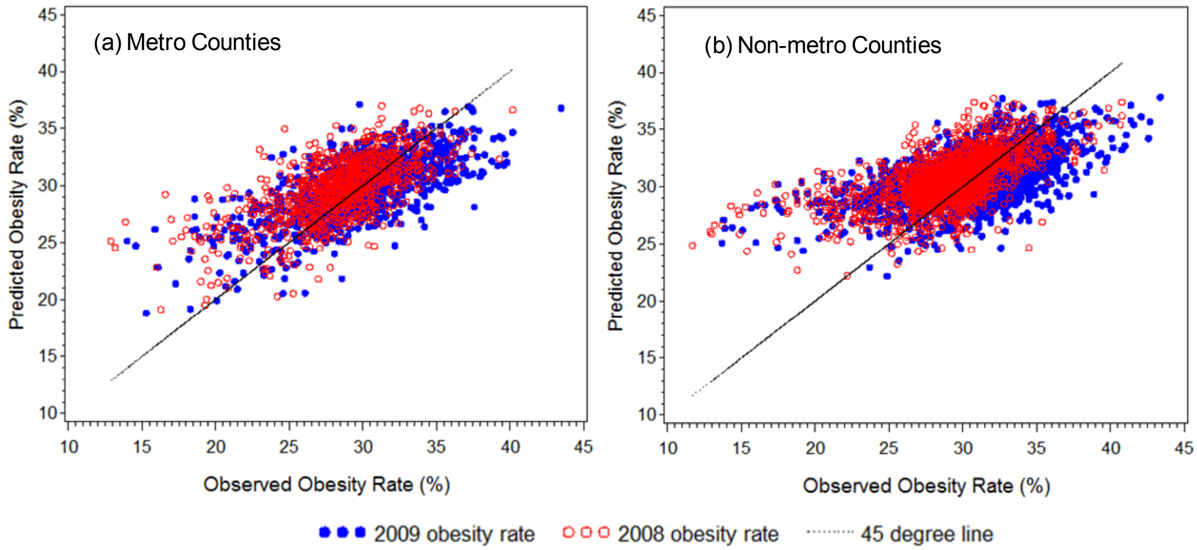


Figure 4-10 Predicted vs observed county level obesity rate for 2008 and 2009.

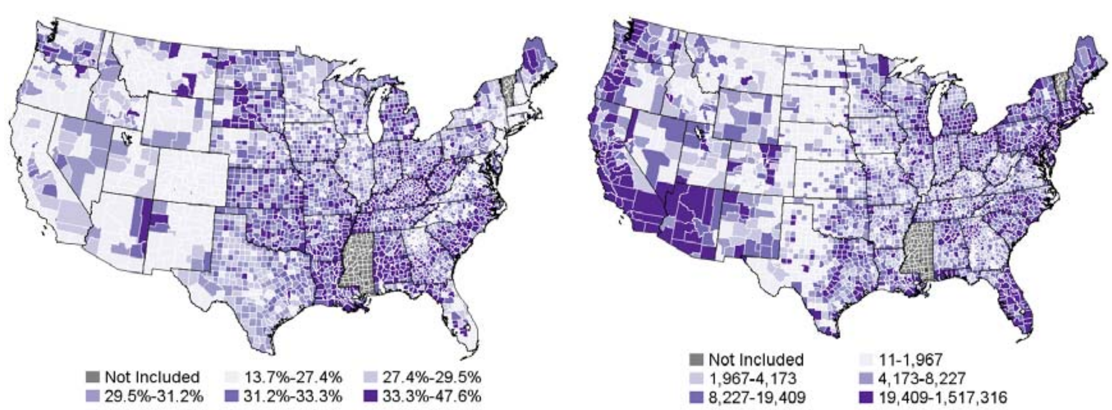
4.3.4 Marginal Effects of Food Retailer Type

In this section we estimated the marginal effect of each food retailer type on obesity rate. Table 4-6 shows estimated percent change in obesity rate with a corresponding 95% confidence interval for a geographic area by adding a single food store for each of the four food retailer types. As mentioned previously, all four food retailer types were significant for metropolitan regions, while for non-metro regions only supercenters and specialty food stores were significant. It is estimated that for a typical metro region, the location of a supercenter would increase obesity by between 26 to 30 persons per 10,000; the location of a grocery story would decrease obesity by between 7 to 8 persons per 10,000; the location of a convenience store would increase obesity by between 4 to 5 persons per 10,000; and the location of specialty food store would decrease obesity by between 25 to 30 persons per 10,000. For a non-metropolitan region, it is estimated that the location of a supercenter would increase obesity by between 24 to 26 persons per 10,000 and the location of specialty food store would decrease obesity rate by between 35 to 38 persons per 10,000.

The marginal effect for each county are shown in Figure 4-12 in terms of obesity rate and Figure 4-13 in terms of number of persons that become obese. In addition, the obesity rate and obesity pupulation are mapped in Figure 4-11. Five response levels were used for each map, where quintiles were used to decide the range for each level. Note that “not included” is for all counties in Mississippi and Vermont as well as those counties for which there were missing values. Since grocery stores and convenience stores were only significant for metropolitan regions, the marginal effect was not estimated for non-metro counties.

Table 4-6 Marginal effect on obesity rate by type of stores.

	Metro (N=1000)		Non-metro (N=1897)	
Type of food store	**Mean**	**95% CI**	**Mean**	**95% CI**
Supercenters	0.28%	(0.26%, 0.30%)	0.25%	(0.24%, 0.26%)
Grocery Stores	-0.08%	(-0.07%, -0.08%)	/	(/ /)
Convenience Stores	0.05%	(0.04%, 0.05%)	/	(/ /)
Specialty Food Stores	-0.27%	(-0.25%, -0.30%)	-0.36%	(-0.35%, -0.38%)



Note: Adults are persons no younger than 20 years old.

Figure 4-11 a) Adults obeisty rate in 2009; b) Number of adult population with obesity.

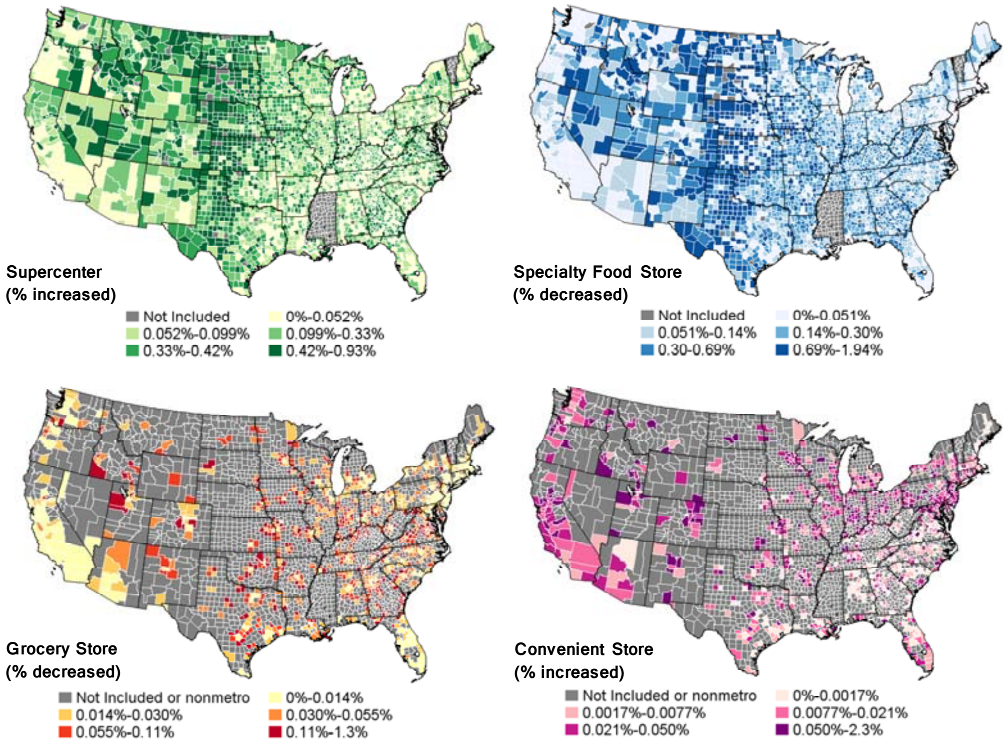


Figure 4-12 Marginal effect on adult obesity rate (%) by type of stores.

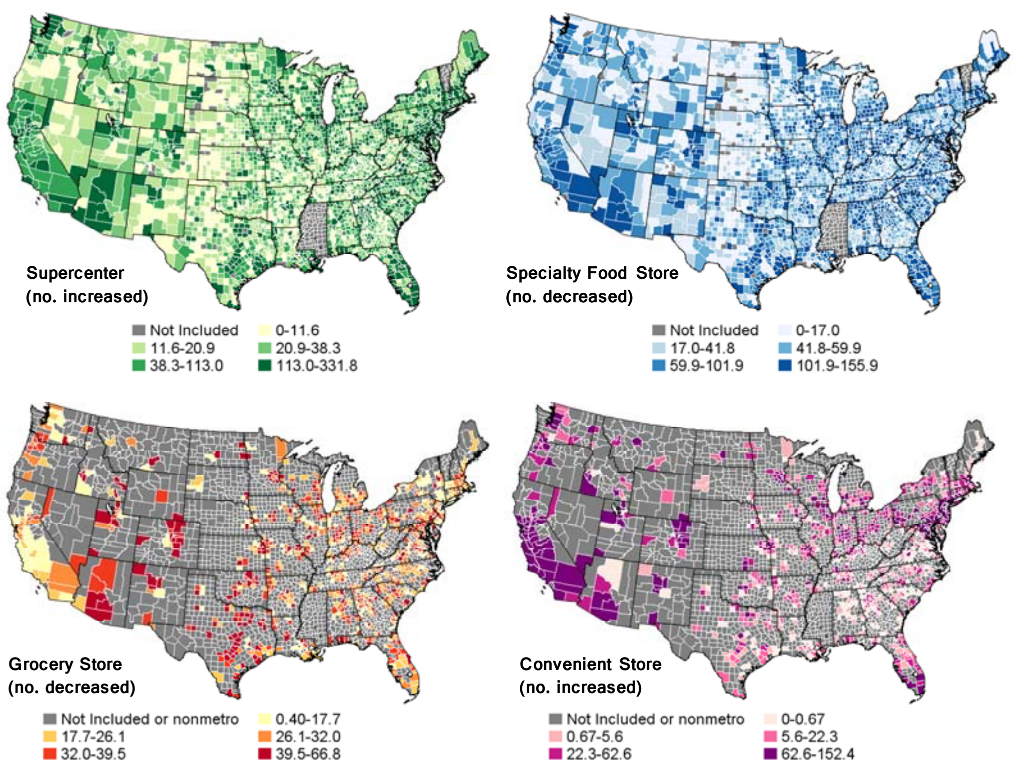


Figure 4-13 Marginal effect on obesity (no. of adult population) by type of stores.

4.4 Discussion

4.4.1 Model Discussion

The nonlinear parametric model shows significantly better fitting accuracy than the linear model from step 1, with 7% and 3% impovementson on adjasted R^2 values for metro and non-metro areas respectively. When compared to the spline regression model in step 2, the R^2 values are worsened by 3 to 4 %. However, the number of coefficients in the model is significantly reduced — the degree of freedom is reduced from 111 to 44 for the metro model, and 116 to 50 for the non-metro model. Further, when looking into the residual by regressor plots, the curventure patterns previously revealed in the linear model are corrected in the nonlinear model.

The advantage of using the spline optimal fitting are that: (1) it allows for flexible shapes, and (2) it achieves an additive model that reduces the interaction among predictors to a minimum through optimally selecting coefficients. The disadvantages are that the achieved transformation may not be a smooth curve and that the nonparametric regression model is hard to assess and apply. The nonlinear parametric model behaved equally as well as the spline model, while overcoming the above shortcomings.

4.4.2 Results Discussion

Several factors are found to be associated with obesity rate from this regression analysis. Generally speaking, obesity was found to be positively associated with lower store accessibility and car ownership, higher poverty rate, higher unemployment rate, higher proportion of whites to non-whites, lower sex ratio of male to female, and a higher level of WIC or SNAP authorized stores. Median age, price ratio of milk to sodas and population density are non-monotonically associated with obesity rate. Food retailer type also had an association with obesity. Specifically, specialty food stores were negatively associated with obesity while supercenters were positively associated with obesity for all areas; convenience stores were positively associated with obesity and grocery stores were negatively associated with obesity in metro areas.

The positive association of supercenter density with obesity is in agreement with Courtemanche and Carden (Courtemanche and Carden, 2011) who used data from the Behavioral Risk Factor Surveillance Systems (BRFSS) matched with Walmart locations and entry data to estimate that an additional supercenter per 100,000 residents increases BMI by 0.24 units. Possible explanations include the offering of cheap food that does not support a healthy lifestyle (Hausman and Leibtag, 2007),the increase of self control problems in the presence of cheaper food (Cutler, Glaeser, and Shapiro, 2003), and lower prices for a wide range of items that lead to a sedentary lifestyle (Basker and Noel, 2009).

For metr regions, the association of obesity rate with specialty food stores and supercenters was in the same direction as for non-metro regions. In addition, grocery stores were found to have a negative association with obesity rate while convenience stores were found to have a positive effect. This is consistent with previous literature (Michimi and Wimberly, 2010; Moore and Roux, 2006; Powell et al., 2007).

There were some differences between metro and non-metro regions. For metro regions, SNAP authorized stores were found to be positively associated with obesity rate. As with the case for WIC authorized stores in non-metro regions, this may be due to the specific targeting of high risk populations.

Although the southeast and midwest regions of the U.S. have the greatest levels of obesity (Control, Prevention, and Atlanta, 2013), they may not necessarily be the ideal region for adopting the food stores intervention strategy. Rather, such intervention strategy should focuse on regions where food stores have a greater impact. Figure 4-12 shows that food store type has the greatest impact in the middle west region of the U.S in terms of the unit improvement in obesity rate. On the other hand, when using reduced obesity population as the measure (Figure 4-13), a higher impact was observed in southwest and east areas. The Reinvestment Fund developed a limited supermarket access measure based on travel distances to reach a supermarket, population density, and car ownership rates (TRF, 2011). The counties that they identified as having limited access match closely with Figure 4-12. However, we feel that the intervention policies shall focus on reduced number of populations instead of the rate, and use Figure 4-13 as a reference guide.

4.5 Applying Marginal Effects for Food Desert Policy

In response to the severity caused by food deserts, several food distribution activities have been initiated and dedicated to solve the problem of obesity. "Let's Move!", promoted by Michele Obama from 2010, is one of these. The initiative is designed to increase a child's intake of fruits, vegetables and whole grains to the recommended level. The Healthy Food Financing Initiative, a partnership between the U.S. Departments of Treasury, Agriculture and Health and Human Services, earmarks over $400 million to provide financing support of developing stores that sell fresh and healthy food in underserved areas. The initiative incentivizes healthy food retailers in the form of tax credits, grants or low-cost loans, and technical assistance. The rather ambitious goal of the initiative is to eliminate food deserts across the country within seven years.

The Healthy Food Financing Initiative was modeled after The Pennsylvania Fresh Food Financing Initiative (FFFI), which is one of the more successful programs. FFFI provides financial support to food retailers to operate in underserved and low-income communities in order to increase the access to nutritious and affordable food. The incentives are designed to help the food retail enterprises overcome the barriers of higher infrastructure costs and credit risks in those economically distressed communities. The program began in 2004 and ended in 2010 when all of the funds were deployed. Over this time period, FFFI committed $73.2 million in loans and $12.1 million in grants to

88 projects, which were expect to bring 1.67 million square feet of fresh food retail space across Pennsylvania. Moreover, thousands of job positions were created through the program. Several grant and loan products were available for the program, and the details varied based on project type, risk, loan terms and social impact. Rather than subsidizing stores that were otherwise economically unfeasible, the program only provided incentives to attract viable supermarkets into underserved areas. The Reinvestment Fund (TRF) was the administrator of the FFFI program.

Although the FFFI program has ended, we may still learn from its experience and help to improve the design of ongoing or future programs. After reviewing the related documents and reports, there are a few shortcomings worth mentioning. These include:

1) The established contract did not depend on performance. As a result, there was no incentive for increasing the level of effort.

2) Grants tended to be preferable to loans and most applicants request the maximum grant amount and did not request a loan at all. Since TRF needed the grant fund pool to last as long as the loan fund, a well-established strategy is needed to allocate grant and loans among projects appropriately.

3) The resulting health benefit from a loan or grant was neither estimated nor well taken in consideration when designing the contract.

In this section, we present a strategy to help address these three issues.

4.5.1 Incentive Contract Design

The first application deals with the case where an operator considers joining a foundation such as the FFFI, and already has a targeted location or area in mind. The foundation would like to design a contract that is attractive enough to the operator but also elicits a desired level of effort. A Principal-Agent framework serves the dual function of risk sharing and encouraging productive work, and is a useful way to formulate the contracting problem facing by the foundation. The objective, in the language of the Principal-Agent problem, is to construct a contract so that the principal's desired action is also the agent's desired action (incentive compatibility). The agent decides whether to accept the risk-sharing contract or financial incentive (participation) in order to maximize his own utility.

Consider the funding program as the principal, who contracts with a single retailer, i.e., the agent. The operator devotes certain effort to the business, which is not directly observable by the principal. We assume such effort has a monetary equivalent, *a*. This effort will lead to an observable outcome of *x*, which has two components: an annual sales of $p(a)=a+\varepsilon$,where ε is normally distributed with a mean of zero and a variance $\sigma_p^{\ 2}$; and expected reduced health costs related to obesity, $h(a)$, resulting from the increased community's access to healthy and affordable foods, where $h(a)$ is a non-decreasing function of effort level. Let's assume that the reduced health costs is of linear form: $h(a)=ma+\delta$, where *m* is selected based on the marginal effect at the target area estimated from the obesity model presented earlier, δ is normally distributed with a mean of zero and a variance $\sigma_h^{\ 2}$. Therefore, the output is defined as:

$$x(a)=p(a)+h(a)=a+ma+\varepsilon+\delta .$$

Further, the principal is assumed to be risk neutral and the agent is assumed to be risk averse. Several reasons could prohibit the retailers from entering the food desert to start new businesses: a lower expected future profit due to the smaller size of the market, higher distribution costs in isolated areas with poor transportation infrastructure, additional training cost from an unskilled labor pool, increased security cost in areas with higher crime rates, as well as the local tax regimes and zoning lows, etc (Besharov, Bitler, and Haider, 2011; Bonanno, 2012). These economic considerations make it more likely that food retailers will not choose to locate in those areas. We therefore assume that the food retailers are risk-averse and have a utility function of

$$U(w)=-e^{-rw},$$

where the coefficient r is the measure of the agent's level of risk aversion, w is the retailer's profit.

Shavel (Shavell, 1979) proved that when agent is risk averse a Pareto optimal contract shall pay the agent depending on his outcome, and will never ask him to bear all of the risk. We therefore use a performance based linear contract of the form:

$$s(x) = \alpha + \beta x .$$

There are three constraints that need to be defined for the optimization problem. From the agent's perspective, he maximizes his utility under a given incentive contract, which is equal to the incentive minus the cost (compatibility constraint). We assume the operator has another market opportunity that he can choose with a utility of $\bar{U}$ (participation constraint). His effort is limited by the demand of the corresponding community (demand constraint). We want to find a Pareto optimal contract by solving the Principal-Agent problem as follows:

$$\underset{s(x)}{\text{Maximze}} \quad E[x(a) - s(x)]$$

s.t. $\quad a \in \arg\max_a E[U(s(x) - c(a))]$

$$E[U(s(x) - c(a))] \geq \bar{U}$$

$$x(a) \leq \bar{d} + h(a)$$

First, we solve the agent's compatibility constraint.

$$\max_a E[U(s,a)] = \max_a E[-e^{-r(s(a)-c(a))}]$$

$$= \max_a - e^{-r(E[s(a)] - \frac{r}{2}\text{var}[s(a)] - c(a))} = \max_a (\alpha + \beta a + \beta m a - \frac{r}{2}\beta^2\sigma^2 - c(a))$$

The first-order condition gives of the solution to the above function as:

$$\beta = \frac{c'(a)}{1+m} .$$

These results simplify the corresponding problem to:

$$\max_{\alpha,\beta} E((1-\beta)(a+ma)-\alpha)$$

$$\text{s.t. } \beta = \frac{c'(a)}{1+m}$$

$$\alpha + \beta a + \beta ma - \frac{(\sigma_p^{\ 2} + \sigma_h^{\ 2})r}{2}\beta^2 - c(a) \geq \bar{u}$$

$$x(a) \leq \bar{d} + h(a),$$

where $\bar{u} = -\ln(-\bar{U})/r$.

As the targeted areas by definition have unmet demand, it is reasonable to assume that the demand constraint will be tight at optimality and can therefore be ignored. As residents in the targeted area are more likely to be obese or overweight and have lower income, supermarkets that sell healthy and fresh foods will need to apply additional effort to achieve a similar level of revenue when compared to equivalent stores in wealthier communities. Such effort might include advertising, organizing special discount events, operating a store bus for areas with low car ownership, etc. We assume a concave cost function of the form $c(a) = ca^2/2$, since compared to a linear cost function that are more generally adopted, the agent shall pay more to achieve the same level of sales.

The maximization problem can be easily solved and results are:

$$\alpha^* = \bar{u} - \frac{(1+m)^4[(1+m)^2 - cr(\sigma_p^{\ 2} + \sigma_h^{\ 2})]}{2c[(1+m)^2 + cr(\sigma_p^{\ 2} + \sigma_h^{\ 2})]^2}$$

$$\beta^* = \frac{(1+m)^2}{(1+m)^2 + cr(\sigma_p^{\ 2} + \sigma_h^{\ 2})}$$

$$a^* = \frac{(1+m)^3}{c(1+m)^2 + c^2 r(\sigma_p^{\ 2} + \sigma_h^{\ 2})},$$

where α^* is the total fixed subsidy paid by the principal, and β^* is the portion of the outcome x belongs to the agent. As the marginal health benefit (m) increases, the total fixed subsidy (α^*) will decrease; while the performance-based subsidy βx will increase.

We use Philadelphia as an example to illustrate the design of an incentive contract. Philadelphia's low-income communities have significant unmet need for fresh food. Consider a risk averse operator ($r = 1$) that is considering joining the market, and has an expected annual profit of over \$5 million per year ($\overline{u} = 5$). According to the TRF Report (Goldstein, 2008), the medium sales from stores within a 5 minute driving time is approximately \$13 million. For this example, we assume the standard deviation is \$1 million ($\sigma_p = 1$) and the cost coefficient is $c = 0.1$. Based on the obesity model, we determine the reduced number of obese individuals when additional grocery store is opened; assume this value is 50. As mentioned previously, the medical costs for people who are obese were \$1,429 higher than those with normal weight (Finkelstein et al., 2009). Let's assume that, a new store with an average effort level of \$13 million will save the obesity related medical cost by approximately $1429 \times 50 = \$71,450$. Thus, $m \approx 0.07145 / 13 = 0.0055$. We also assume that the standard deviation of reduced medical cost is \$10,000 ($\sigma_h = 0.01$). Solving the principal-agent problem, we have the optimal incentive contract as,

$$\alpha^* = 1.228$$

$$\beta^* = 0.910$$

$$a^* = 9.150.$$

Consider \$1.228 million as the fixed total subsidy provided to the operator at the beginning of the business and is independent of performance (e.g. in the forms of grant and loan). The other component is performance-based: 91.0% of the sales outcome plus reduced obesity related medical costs will belong to the agent, where the latter is an award for the store's contribution to reducing the local obesity rate. A profit driven operator will make the effort to achieve the sales goal of \$9.15 million. The expected annual profit of the operator and the funding program are respectively,

$$E(s(x)-c(a)) = \$5.0 \text{ million}$$

$$E(x(a)-s(x)) = -\$0.4 \text{ million} .$$

By designing an incentive contract in this manner, the foundation and operator are sharing the risk in a way that will maximize the benefit for both parties. Meanwhile, the agent will be encouraged to work hard and the program will be sustainable. The traditional method of a fixed subsidy would leave the operator with no incentive to take effort. In this specific case of opening stores in a distressed area, the operators are more likely to be risk averse. By designing an outcome based contract, the agent will receive an outcome-based payment but he would never bear all the risk.

In the above model, we assumed that the health related outcome $h(a)$ is in a linear form. Let's now assume give an example to show diminishing returns. A nonlinear scenario where $h(a)$ is increasing and strictly concave with respect to a. This assumption is in accordance with the law of diminishing marginal utility, which means that the marginal health benefit decreases as the units of effort increase. We assume $h(a)$ is in the form of $k(1\text{-}e^{-a})$. Similar to the above example, it is assumed that an effort level with a target sale of \$13 million is expected to reduce obesity related medical cost by \$71,450. Therefore, $h(a)=0.0714502(1\text{-}e^{-a})$. By applying first-condition, the compatibility constrain is updated as,

$$\beta = \frac{ca}{1+0.0714502e^{-a}} \ .$$

Now we may easily solved the nonlinear optimization problem in LINGO. The optimal solution is: $a^{*}=9.091$, $\alpha^{*}=1.216$, $\beta^{*}=0.909$. The expected outcome is $E(x)=a+h(a)=9.162$ million. If we double this optimal effort, then $E(x)$ will become \$18.254 million. Note that this is less than two times of the outcome at optimal effort level. In other words, a diminishing return is observed in this case.

4.5.2 Grant and Loan Division

The second application of the obesity model is where one operator considers locating in one of multiple locations. From the incentive contract design, we already have the total subsidy. The problem, however, is how to divide the total fixed subsidies into loans and grants. Suppose a retailer that is considering joining the foundation, but has not made the decision of where to establish the new store. As mentioned earlier for the FFFI program, the majority of applicants preferred grants to loans and nearly all request the maximum grant amount. It is therefore reasonable to assume that a higher portion of grants would be more attractive to the applicant. TRF, the administrator of FFFI program, did not give a clear explanation on their strategy of grant-loan allocation. They only mentioned that the approval decision was based on the extent to which the project fit within TRF's mission, as well as the long-term viability of the proposed business. However, the health benefit was not well considered when making the decision.

From the principal-agent framework, we have the total fixed subsidy α_i for project i. Assume the proportion of the grant in the total subsidies is g_i, and $(1-g_i)$ is the proportion of loan interest subsidy, which is the market interest rate minus the total interest paid at terms offered by the funding program. As a higher value of g_i is preferred by the agent, we could design the contract purposefully to attract the operators to communities where higher health benefits can be gained from a new store. From the obesity model, we have the marginal effect of reduced obesity rate, m_i. We assign a value of t_i proportional to m_i, i.e., $g_i = k \cdot m_i$. Therefore, we can formulate the problem as,

$$\sum_i \alpha_i \cdot g_i = k \sum_i \alpha_i \cdot m_i = \text{Grant Budget}$$

Solving for k, we have $k = \dfrac{\text{Grant Budget}}{\sum_i \alpha_i \cdot m_i}$, and the operator chooses location i and will receive km_i portion of the total fixed subsidy in a grant form.

4.5.3 Strategic Planning

To this point, we have only considered design of contract. The third application assists local government to strategically design a food store establishment plan given a reduction objective of obesity and various intervention options being available. The index and parameters are listed as follows:

i=1,2,3. type of store: 1. supermarket or grocery stores (metro only); 2. specialty food stores; 3 for convenience stores

j=1,2,3. type of interventions: 1. upgrade convenience store without fuel pump to grocery store; 2. introduce new specialty food store; 3. introduce new supermarket or grocery store.

s_i: current number of stores of type i

x_i: increased number for store type i (decisional variable). Note that both intervention 1 and 3 will increase the number of type 1 stores (i.e. supermarket or grocery stores)

α_j: associated fixed subsidy per store for intervention type j (estimated from Principal-Agent framework)

$t(s_1, s_2, s_3)$ is the obesity rate given the number of stores by type. $t'(s_i)$ is the Box-Cox transformed obesity rate. $f_i(s_i)$ is store type i 's contribution to the transformed obesity rate $t'(s_i)$. According to our obesity model, all $f_i(x_i)$ are sigmoid functions in the form:

$$f_i(s_i) = a_i + b_i \frac{1}{1+e^{-(s_i-\lambda_i)\alpha_i}}$$

where the variables are retailer type dependent.

The objective is to minimize the financial investment given a target reduction rate of obesity, $\overline{z}$. The optimization problem is formulated as:

$$\min \alpha_1(-x_3) + \alpha_2 x_2 + \alpha_3(x_1 + x_3)$$

s.t. $t'(s_1 + x_1, s_2, s_3) = t'(s_1, s_2, s_3) + f_1(s_1 + x_1) - f_1(s_1)$

$$t'(s_1, s_2 + x_2, s_3) = t'(s_1, s_2, s_3) + f_2(s_2 + x_2) - f_2(s_2)$$

$$t'(s_1, s_2, s_3 + x_3) = t'(s_1, s_2, s_3) + f_3(s_3 + x_3) - f_3(s_3)$$

$$3t(s_1, s_2, s_3) - [t(s_1 + x_1, s_2, s_3) + t(s_1, s_2 + x_2, s_3) + t(s_1, s_2, s_3 + x_3)] \geq \overline{z}$$

$x_1 + x_3 \geq 0$ upgraded stores through option 3 is positive

$-x_3 \leq \overline{s}$ upgraded convenience store is limited by current number of no-fuel pump stores

$x_1, x_2 \geq 0, \;\; x_3 \leq 0, \;\;$ integers

We use Richmond County in Georgia to illustrate this model. Richmond is geographically the largest county in GA with a population of 199,768 and 141,502 are older than 20, among whom 32.5% are obese ($t(s_1, s_2, s_3) = 0.325$). In 2009, Richmond County had 37 grocery stores and supermarkets ($s_1 = 37$), 12 specialty food stores ($s_2 = 12$), 10 convenience stores without fuel pumps ($\overline{s} = 10$) and 106 with fuel pumps ($s_3 = 106$). The fixed subsidy can be estimated from solving the Principal-Agent formulation. Assume the total annual fixed subsidies for each type of intervention are: \$0.1 million for upgrading a convenience store without fuel pumps to a grocery store (α_1=0.1), \$0.11 million for opening a new specialty food store (α_2=0.11), and \$0.3 million for opening a new grocery store (α_3=0.3). Assume the target is to reduce the obesity rate by 1% ($\overline{z} = 0.01$). This would imply 1,415 adults would no longer be obese from the intervention. The functions of $f_i(s_i)$ can be found from Table 4-5.

By solving the above nonlinear integer programming problem using LINGO, the global optimal solution for x_i are [10, 21, -10], i.e., to upgrade 10 convenience stores to grocery stores and establish 21 specialty food stores. The total fixed subsidy per year is roughly \$3.31 million. It is interesting to note from this result that instead of the traditional approach of opening new grocery stores, our model shows that upgrading current convenience stores and encouraging the development of specialty food stores would be more cost efficient in this case.

4.6 Conclusion

Food retailer type was found to be associated with obesity rate even after controlling for food environment, food assistance programs, and socio-demographic characteristics. In particular, the study showed that in metropolitan areas obesity rate was positively associated with supercenters and convenience stores and negatively associated with grocery stores and specialty food stores. In non-metropolitan areas it was found that obesity rate was positively associated with supercenters and negatively associated with specialty food stores. The marginal effect of the introduction of a food retailer type in geographic region was also estimated.

The marginal effect measures can be useful in designing effective interventions based on food retailer type. Several organizations such as the U.S. Department of Agriculture are encouraging the introduction of grocery stores and specialty food stores. These measures would help to determine the best allocation of limited financial resources to reduce obesity levels. We are now working on the improvement of the model by including spatial correlation.

Chapter 5

Conclusions

Increasing pressure from limited health care resources and low quality of care requires decision makers to design intervention strategies in a more cost-effective way. As such, cost-effective modeling is becoming increasingly popular in public health studies With various options of cost-effective models available, the selection of the right approach should be based on the study objective and the nature of data.

Markov model is a better choice for cases where the researcher concerns with multiple health states which may happen at anytime and multiple-times during the studied time frame. The transitions from one state to another will be estimated. Instead of using the static measures such as utilization, the time dependent transition probability would be used for assessment instead. This is especially useful when activities in the previous time cycle is of great interest to the decision maker.

Computer-based simulation is a great tool for cases where the relationships between health outcome and its influential factors are relatively complicated and the data is insufficient to support for a traditional statistical modeling. Such method helps to understand the complex interactions among factors within the system, and more importantly, helps to assess the potential performance of intervention strategies before they are put into practices. A set of assumptions shall be made for the simulation, which are drawn from the historical data though. For example, in the Tuberculosis intervention study, the target hospital would need to provide the distribution of patients arrival rate, staying time, health-care working schedule, etc. The simulation will keep track of the events and outcome for a period of time, and the decision maker may then summarize the statistics and do the trade-off among different intervention options.

Complex statistical modeling is an extension of the standard statistical modeling techniques such as multiple regression. Statistical regression is applied extensively by health care researchers to filter the factors that significantly affect the health outcome, and do the prediction accordingly. However, such models may less of prediction accuracy in many cases, and when logistic regression is applied, the odds ratio instead of real outcomes would be estimated instead. To overcome those deficiencies, more advanced statistical modeling technique with higher prediction accuracy is recommended in the health care decision making.

. In this dissertation, three frameworks were illustrated using cost-effective modeling techniques discussed above to solve the public health problems.

The first problem concerned the evaluation of the dental care system for children. We developed a stochastic model using longitudinal data to assess the appropriateness of care using publicly available data. This kind of data, although providing less detail on procedures delivered as compared to clinical claims data, represents a more closed system with information collected from all potential dental providers. For this reason, longitudinal data like MEPS are more reliable when drawing nationally representative conclusions. By linking the service mix from MEPS data to the four stages of care, i.e. Initial, Maintenance, Episodic and Nonuse, a Markov chain model was then applied and annual transition probabilities across stages of care were estimated accordingly. Such recurring patterns provide the evidence base for how well patients are integrated into the dental system and where the problem areas are. Compared to private-insured children, Medicaid and non-insured children were less likely to enter the system and once entered, were less likely to stay in the system for further preventive treatment. This issue is primarily due to two factors: i) from the demand side, Medicaid and non-insured children come from families under greater financial pressure and they usually have less knowledge about the benefits of preventive services, and ii) from the supply side, the low Medicaid reimbursements for dental services relative to private market rates could be a key contributor. Meanwhile, as the private insured group outperforms the others, it could potentially serve as a benchmark for dental care programs, which is indeed a missing element currently.

Our results were validated based on the similarity in patterns as compared an earlier study that used clinical claims data with readily defined stages of care. The key difference in the results is that patients in our analysis were more likely to transition to maintenance and less likely to exit the system. In addition to the differences in time and population constitution, another possible reason is the misclassification of I and M. Since MEPS does not specify whether an examination was comprehensive or periodic, we had to use cut points based on costs to differentiate between these two. It is our recommendation that longitudinal data to include such information in the future, which will greatly improve the accuracy of the results. Another advantage of this model is its predictive ability. For example, we may predict the stage distribution of patients at the stable state, as well as the probability that a patient takes a specific path. We also illustrated the possibility of using this model to estimate the government investment required to reach a specific quality target.

The second problem deals with the intervention of Tuberculosis transmission in health care settings. In this study, we developed an infection risk model based on both biological and physical principles while taking into consideration the primary intervention measures. The simulation-based model serves as the basis for assessing the efficiency of various intervention strategies, and assists the selection of the most cost-effective intervention option for a clinic representative that found in a resource-constrained country based on CDC specified data. Using the simulation results on expected yearly infection cases, the hospital may decide the best combination of devices available in the market, the isolation strategy to protect HIV+ patients from being infected, and how frequent health care workers shall be tested for TB infection. In addition, the cost issues of applying different intervention strategies are considered.

A distributive model with space segmentation was developed and studied, and provides strong evidence that our hypothesis that susceptible patients sitting closer to the infectious sources will have a greater risk of being infected. This model will be especially useful for resource-restricted areas where building physical isolation room is not practical. Instead, such areas may consider using virtual isolation in one general waiting room, to separate patients with TB symptoms from others without such symptoms. So far, we have studied two and four segmentation models. Depending on the room dimension and the level of concerns, the clinical setting may increase the number of segmentations to meet their specific needs. We are currently collaborating with the CDC to gather experimental data from hospitals such that validity of the risk model may be further assessed by practice. This will be completed in future work.

The third problem quantifies the impact of food retailers type on obesity prevalence in geographic regions (counties). A three-step nonlinear parametric model was developed using publicly available data, and the marginal effects on obesity rate from the addition of a new food retailer type is estimated for each county. Several factors were found to be significantly associated with obesity rate, including lower store accessibility and car ownership, higher poverty rate, higher unemployment rate, higher proportion of whites to non-whites, lower gender ratio of male to female, and a higher level of WIC or SNAP authorized stores. More importantly, it is found that obesity rate is positively associated with supercenters and negatively associated with specialty food stores in both metro and non-metro areas. For metro areas, obesity rate is also positively associated to convenience stores are negatively associated with grocery stores. The marginal effect is very useful in identifying regions where interventions based on food retailer type would be most effective.

There are several advantages of our study over the previous ones. First and most import one is quantifying the impact accurately. The adjusted R^2 is improved from the best of 0.06 in previous studies to 0.38-0.49 in our study. Second, while the majority of previous studies have focused on a selective local area, we include a national analysis at the county level and thus provided more representative insights to the obesity problem. Third, our model is more comprehensive than those in previous studies. All major types of food retailers are included and are adjusted by other factors, such as, prices, food assistance coverage, and socioeconomic variables.

We presented three possible applications of our model to the design of intervention policies, including: an incentive contract depending on both sales and health outcomes, that maximizes the foundation's utility and elicits retailers' effort to a desired level, proper allocation of grants to attract potential retailers into areas where higher obesity reduction is expected, and an optimal store establishment plan that minimizes the financial investment while achieving the obesity reduction goal. All these designs have taken health outcomes into proper consideration, a missing element in the current policy design framework.

Several limitations to this study should be mentioned, which are also possible directions for future work to improve the model. First, information on distance to store and foot square of fresh food were not available. We used store density and proportion of population with low accessibility and no cars as a proxy. Second, it was assumed that individuals only shop in a local geographic area and do not cross regions. Third, restaurants, another important component of food environment, were not included in the model.

This dissertation contributed to the invention policy design of three totally different public health areas and illustrated the great advantages of applying cost-effective modeling into the policy design, decision and assessment processes. Such techniques maximize benefits using limited public resources based on sound data-based evidences, and should be popularized into more areas of public health decision making.

References

ADA. 2001.*Future of dentistry: today's vision, tomorrow's reality*. American Dental Association. Health Policy Resources Center.

ADA. 2009.*2009 Survey of Dental Fees*. Chicago, IL: American Dental Association (ADA).

Agency for Healthcare Research and Quality. 2012. Table 2.2: Percent of Children Age 2 - 17 with a Dental Visit in the Past Year: United States, 2009. Available at http://meps.ahrq.gov/mepsweb/data_stats/quick_tables.jsp=.

American Dental Association Council on Scientific Affairs. 2006. Professionally applied topical fluoride Evidence-based clinical recommendations. *The Journal of the American Dental Association* 137: 1151-9.

Azarpazhooh, A., and P. Main. 2009. Efficacy of dental prophylaxis (rubber cup) for the prevention of caries and gingivitis: a systematic review of literature. *British dental journal* 207: E14-E.

Baicker, K., and A. Chandra. 2004. Medicare Spending, The Physician Workforce, and Beneficiarie'Quality of Care.

Basker, E., and M. Noel. 2009. The Evolving Food Chain: Competitive Effects of Wal‐Mart's Entry into the Supermarket Industry. *Journal of Economics & Management Strategy* 18: 977-1009.

Berwick, D.M., T.W. Nolan, and J. Whittington. 2008. The triple aim: care, health, and cost. *Health Affairs* 27: 759-69.

Besharov, D.J., M. Bitler, and S.J. Haider. 2011. An economic view of food deserts in the United States. *Journal of Policy Analysis and Management* 30: 153-76.

Block, D., and J. Kouba. 2006. A comparison of the availability and affordability of a market basket in two communities in the Chicago area. *Public health nutrition* 9: 837-45.

Bonanno, A. 2012. Food Deserts: Demand, Supply, and Economic Theory. *Choices* 27.

Box, G.E., and D.R. Cox. 1964. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*: 211-52.

Brown, A.F., R.B. Vargas, A. Ang, and A.R. Pebley. 2008. The neighborhood food resource environment and the health of residents with chronic conditions. *Journal of general internal medicine* 23: 1137-44.

Caro, J.J., J. Möller, and D. Getsios. 2010. Discrete event simulation: the preferred technique for health economic evaluations? *Value in health* 13: 1056-60.

Caspi, C.E., G. Sorensen, S. Subramanian, and I. Kawachi. 2012. The local food environment and diet: a systematic review. *Health & Place* 18: 1172-87.

CBO. 2013.*The Budget and Economic Outlook: Fiscal Years 2013 to 2023*.

CDC. 1994.*Guidelines for preventing the transmission of Mycobacterium tuberculosis in health-care facilities, 1994*. US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention.

CDC. 2005.*Guidelines for preventing the transmission of Mycobacterium tuberculosis in health-care settings, 2005*: US Department of Health and Human Services, Public Health Service, Centers for Disease Control and Prevention.

Chen, S., R.J. Florax, S. Snyder, and C.C. Miller. 2010a. Obesity and access to chain grocers. *Economic Geography* 86: 431-52.

Chen, S., R.J.G.M. Florax, S. Snyder, and C.C. Miller. 2010b. Obesity and access to chain grocers. *Economic Geography* 86: 431-52.

Chen, Z., S.T. Yen, and D.B. Eastwood. 2005. Effects of food stamp participation on body weight and obesity. *American Journal of Agricultural Economics* 87: 1167-73.

Clarke, G., H. Eyre, and C. Guy. 2002. Deriving indicators of access to food retail provision in British cities: studies of Cardiff, Leeds and Bradford. *Urban Studies* 39: 2041-60.

Congressional Budget Office. 2012. The 2012 Long-Term Budget Outlook. Available at http://cbo.gov/publication/43288 (accessed June, 2012).

Control, C.f.D., Prevention, and G. Atlanta. 2013.*Overweight and obesity. Adult obesity facts*.

Courtemanche, C., and A. Carden. 2011. Supersizing supercenters? The impact of Walmart Supercenters on body mass index and obesity. *Journal of Urban Economics* 69: 165-81.

Cummins, S., and S. Macintyre. 2002. "Food deserts"—evidence and assumption in health policy making. *BMJ* 325: 436-8.

Cutler, D., E. Glaeser, and J. Shapiro. 2003.*Why have Americans become more obese?* : National Bureau of Economic Research.

De Boor, C. 1978. A practical guide to splines.

Decker, S.L. 2011. Medicaid payment levels to dentists and access to dental care among children and adolescents. *JAMA: the journal of the American Medical Association* 306: 187.

Delta Dental. Utilization Review Guidelines, January 1, 2013. Available at https://www.deltadentalrionline.com/Public/PDF/DDRI_ProviderManual.pdf (accessed September 30, 2013.

Department of Health and Human Services. 2013.*Patient Protection and Affordable Care Act; standards related to essential health benefits, actuarial value and accreditation. Final rule.* . 12833-72.: Fed Regist.

Devlin, N., and J. Sussex. 2011. Incorporating multiple criteria in HTA: methods and processes. *London: Office of health economics*.

Dolan, P., R. Metcalfe, V. Munro, and M. Christensen. 2008. Valuing lives and life years: anomalies, implications, and an alternative. *Health economics, policy, and law* 3: 277.

Drew Ambulatory Review Team (DART). 1976.*Standards for dental services*. Charles R. Drew Postgraduate Medical School, Los Angeles, CA.

Dye, B.A., X. Li, and G. Thornton-Evans. 2012. Oral Health Disparities as Determined by Selected Healthy People 2020 Oral Health Objectives for the United States, 2009-2010. *NCHS data brief*: 1.

Eames, I., J. Tang, Y. Li, and P. Wilson. 2009. Airborne transmission of disease in hospitals. *Journal of The Royal Society Interface* 6: S697-S702.

Eisenhauer, E. 2001. In poor health: Supermarket redlining and urban nutrition. *GeoJournal* 53: 125-33.

Emanuel, E.J. 2012. Where are the health care cost savings? *JAMA: the journal of the American Medical Association* 307: 39-40.

Evans, A.E., R. Jennings, A.W. Smiley, J.L. Medina, S.V. Sharma, R. Rutledge, M.H. Stigler, and D.M. Hoelscher. 2012. Introduction of farm stands in low-income communities increases fruit and vegetable among community residents. *Health & Place* 18: 1137-43.

Evans, G., N. Wells, M. Schamberg, L. Dubé, A. Bechara, and A. Dagher. 2010. The role of the environment in socio-economic status and obesity. *Obesity prevention: The role of the brain and society on individual behavior*: 713-25.

Finkelstein, E.A., J.G. Trogdon, J.W. Cohen, and W. Dietz. 2009. Annual medical spending attributable to obesity: payer-and service-specific estimates. *Health Affairs* 28: w822-w31.

Flegal, K.M., M.D. Carroll, C.L. Ogden, and C.L. Johnson. 2002. Prevalence and trends in obesity among US adults, 1999-2000. *JAMA: the journal of the American Medical Association* 288: 1723-7.

Fontaine, K., and I. Barofsky. 2001. Obesity and health‚Äêrelated quality of life. *Obesity reviews* 2: 173-82.

Franco, M., A.V. Diez Roux, T.A. Glass, B. Caballero, and F.L. Brancati. 2008. Neighborhood characteristics and availability of healthy foods in Baltimore. *American journal of preventive medicine* 35: 561-7.

Freed, J.R., M. Marcus, and A.B. Forsythe. 1979. A Markovian model for evaluating dental care programs. *Community dentistry and oral epidemiology* 7: 25-9.

Galvez, M.P., L. Hong, E. Choi, L. Liao, J. Godbold, and B. Brenner. 2009. Childhood Obesity and Neighborhood Food Store Availability in an Inner City Community. *Academic pediatrics* 9: 339.

Gammaitoni, L., and M.C. Nucci. 1997. Using a mathematical model to evaluate the efficacy of TB control measures. *Emerging infectious diseases* 3: 335.

Gibson, D. 2003. Food stamp program participation is positively related to obesity in low income women. *The Journal of nutrition* 133: 2225-31.

Goldstein, I. 2008.*CDFI Financing of Supermarkets in Underserved Communities: A Case Study*: The Reinvestment Fund.

Hagan, J.F., J.S. Shaw, and P.M. Duncan. 2008.*Bright futures: Guidelines for health supervision of infants, children, and adolescents*: American Academy of Pediatrics Elk Grove Village, IL.

Handy, S.L., and K.J. Clifton. 2001. Local shopping as a strategy for reducing automobile travel. *Transportation* 28: 317-46.

Hausman, J., and E. Leibtag. 2007. Consumer benefits from increased competition in shopping outlets: Measuring the effect of Wal‐Mart. *Journal of Applied Econometrics* 22: 1157-77.

Institute of Medicine . Committee on Quality of Health Care in America. 2001.*Crossing the quality chasm: A new health system for the 21st century*: National Academies Press.

International Federation of Health Plans. 2013.*2012 Comparative Price Report: Variation in Medical and Hospital Prices by Country*.

Jago, R., T. Baranowski, J.C. Baranowski, K.W. Cullen, and D. Thompson. 2007. Distance to food stores & adolescent male fruit and vegetable consumption: mediation effects. *International Journal of Behavioral Nutrition and Physical Activity* 4: 35.

Jetter, K.M., and D.L. Cassady. 2006. The availability and cost of healthier food alternatives. *American journal of preventive medicine* 30: 38-44.

Just, D.R. 2006. Behavioral economics, food assistance, and obesity. *Agricultural and resource economics review* 35: 209.

Kaiser Family Foundation. 2012. Health Care Costs: A Primer. Key Information on Health Care Costs and Their Impact. Available at http://www.kff.org/insurance/upload/7670-03.pdf (accessed June, 2012).

Kenyon, T.A., R. Ridzon, R. Luskin-Hawk, C. Schultz, W.S. Paul, S.E. Valway, I.M. Onorato, and K. Castro. 1997. A nosocomial outbreak of multidrug-resistant tuberculosis. *Annals of internal medicine* 127: 32.

Klevens, R.M., J.R. Edwards, C.L. Richards, T.C. Horan, R.P. Gaynes, D.A. Pollock, and D.M. Cardo. 2007. Estimating health care-associated infections and deaths in US hospitals, 2002. *Public health reports* 122: 160.

Ko, G., H.A. Burge, E.A. Nardell, and K.M. Thompson. 2001. Estimation of tuberculosis risk and incidence under upper room ultraviolet germicidal irradiation in a waiting room in a hypothetical scenario. *Risk Analysis* 21: 657-74.

Ko, G., K.M. Thompson, and E.A. Nardell. 2004. Estimation of tuberculosis risk on a commercial airliner. *Risk Analysis* 24: 379-88.

Krukowski, R.A., D.S. West, J. Harvey-Berino, and T.E. Prewitt. 2010. Neighborhood impact on healthy food availability and pricing in food stores. *Journal of community health* 35: 315-20.

Larson, N.I., M.T. Story, and M.C. Nelson. 2009. Neighborhood environments: disparities in access to healthy foods in the US. *American journal of preventive medicine* 36: 74-81. e10.

Lewis, C., A.S. Robertson, and S. Phelps. 2005. Unmet dental care needs among children with special health care needs: implications for the medical home. *Pediatrics* 116: e426-e31.

Liese, A.D., K.E. Weis, D. Pluto, E. Smith, and A. Lawson. 2007. Food store types, availability, and cost of foods in a rural environment. *Journal of the American Dietetic Association* 107: 1916-23.

Marcus, M., A. Koch, and J. Gershen. 1979. A record review model for assessing dental practices. *CDA journal California Dental Association* 7: 51.

Marsh, K., C.J. Phillips, R. Fordham, E. Bertranou, and J. Hale. 2012. Estimating cost-effectiveness in public health: a summary of modelling and valuation methods. *Health economics review* 2: 1-6.

McGinnis, J.M., and W.H. Foege. 1993. Actual causes of death in the United States. *JAMA: the journal of the American Medical Association* 270: 2207-12.

Menzies, D., A. Fanning, L. Yuan, and J.M. FitzGerald. 2000. Hospital ventilation and risk for tuberculous infection in canadian health care workers. Canadian Collaborative Group in Nosocomial Transmission of TB. *Ann Intern Med* 133: 779-89.

Michimi, A., and M.C. Wimberly. 2010. Associations of supermarket accessibility with obesity and fruit and vegetable consumption in the conterminous United States. *International Journal of Health Geographics* 9: 94-112.

Monsivais, P., and A. Drewnowski. 2007. The rising cost of low-energy-density foods. *Journal of the American Dietetic Association* 107: 2071-6.

Moore, L.V., and A.V.D. Roux. 2006. Associations of neighborhood characteristics with the location and type of food stores. *Journal Information* 96.

Moore, L.V., A.V.D. Roux, J.A. Nettleton, and D.R. Jacobs. 2008. Associations of the Local Food Environment with Diet Quality - A Comparison of Assessments based on Surveys and Geographic Information Systems The Multi-Ethnic Study of Atherosclerosis. *American Journal of Epidemiology* 167: 917-24.

Morland, K., A.V. Diez Roux, and S. Wing. 2006. Supermarkets, other food stores, and obesity: the atherosclerosis risk in communities study. *American journal of preventive medicine* 30: 333-9.

Morland, K., S. Wing, A. Diez Roux, and C. Poole. 2002. Neighborhood characteristics associated with the location of food stores and food service places. *American journal of preventive medicine* 22: 23-9.

Morland, K.B., and K.R. Evenson. 2009. Obesity prevalence and the local food environment. *Health and Place* 15: 491-5.

Morton, L.W., and T.C. Blanchard. 2007. Starved for access: life in rural America's food deserts. *Rural Realities* 1: 1-10.

Nardell, E.A. 1993. Environmental control of tuberculosis. *The Medical clinics of North America* 77: 1315.

National Center for Health Statistics. 2012.*Health, United States, 2011: With Special Feature on Socioeconomic Status and Health*. Hyattsville, MD.

National Quality Forum. 2012.*Oral health performance measurement: Environmental scan, gap analysis and measure topics prioritization—Technical report*. 112 pp. Washington, DC: National Quality Forum.

Newacheck, P.W., D.C. Hughes, Y.-Y. Hung, S. Wong, and J.J. Stoddard. 2000. The unmet health needs of America's children. *Pediatrics* 105: 989-97.

Nguyen, D.M., and H.B. El-Serag. 2010. The epidemiology of obesity. *Gastroenterology Clinics of North America* 39: 1.

NIH. 1998.*Clinical guidelines on the identification, evaluation, and treatment of overweight and obesity in adults: The evidence report*. Bethesda, MD: National Institute of Health (NIH).

Office of the Legislative Counsel. 2010. Compilation of Patient Protection and Affordable Care Act

Ogden, C.L., M.D. Carroll, B.K. Kit, and K.M. Flegal. 2012.*Prevalence of obesity in the United States, 2009-2010*: US Department of Health and Human Services, Centers for Disease Control and Prevention, National Center for Health Statistics.

Pearce, J., T. Blakely, K. Witten, and P. Bartie. 2007. Neighborhood deprivation and access to fast-food retailing: a national study. *American journal of preventive medicine* 32: 375-82.

Pinheiro, J.C., and D.M. Bates. 1995. Approximations to the log-likelihood function in the nonlinear mixed-effects model. *Journal of Computational and Graphical Statistics* 4: 12-35.

Powell, L.M., E. Han, and F.J. Chaloupka. 2010. Economic contextual factors, food consumption, and obesity among US adolescents. *The Journal of nutrition* 140: 1175-80.

Powell, L.M., S. Slater, D. Mirtcheva, Y. Bao, and F.J. Chaloupka. 2007. Food store availability and neighborhood characteristics in the United States. *Preventive medicine* 44: 189-95.

Pub L No. 110-224. 2008. Food, Conservation, and Energy Act of 2008.

Riley, E., G. Murphy, and R. Riley. 1978. Airborne spread of measles in a suburban elementary school. *American Journal of Epidemiology* 107: 421-32.

Scott, R.D., and R. Douglas. 2009. The direct medical costs of healthcare-associated infections in US hospitals and the benefits of prevention.

Sharkey, J.R., and S. Horel. 2008. Neighborhood socioeconomic deprivation and minority composition are associated with better potential spatial access to the ground-truthed food environment in a large rural area. *The Journal of nutrition* 138: 620-7.

Shavell, S. 1979. Risk sharing and incentives in the principal and agent relationship. *The Bell Journal of Economics*: 55-73.

Smith, R.D. 2000. The discrete-choice willingness-to-pay question format in health economics: should we adopt environmental guidelines? *Medical Decision Making* 20: 194-204.

Swallen, K.C., E.N. Reither, S.A. Haas, and A.M. Meier. 2005. Overweight, obesity, and health-related quality of life among adolescents: the National Longitudinal Study of Adolescent Health. *Pediatrics* 115: 340-7.

Townsend, M.S., J. Peerson, B. Love, C. Achterberg, and S.P. Murphy. 2001. Food insecurity is positively related to overweight in women. *The Journal of nutrition* 131: 1738-45.

TRF. 2011.*Identifying optimal areas for supermarket development. In: Financing healthy food options: implementation handbook*: The Reinvestment Fund.

van Rijckeveorsel, J. 1982. Canonical Analysis with B-splines. *COMPUSTAT*.

Walker, R.E., C.R. Keane, and J.G. Burke. 2010. Disparities and access to healthy food in the United States: a review of food deserts literature. *Health & Place* 16: 876-84.

Wells, W.F., and W. Stone. 1934. ON AIR-BORNE INFECTION STUDY III. VIABILITY OF DROPLET NUCLEI INFECTION. *American Journal of Epidemiology* 20: 619-27.

Wennberg, J.E., S. Brownlee, E.S. Fisher, J.S. Skinner, and J.N. Weinstein. 2008.*An agenda for change: improving quality and curbing health care spending: opportunities for the congress and the Obama Administration*: Dartmouth Institute for Health Policy and Clinical Practice.

WHO. 2013. *World Health Statistics. May 2013*.

WHO. *Life expectancy by country*. Global Helath Observatory.

WHO. 2006.*The Global Plan to Stop TB 2006–2015: Action for Life Toward a World Free of Tuberculosis.* Geneva: World Health Organization (WHO).

Winsberg, S., and J.O. Ramsay. 1980. Monotonic transformations to additivity using splines. *Biometrika* 67: 669-74.

VITA

Renfei Yan

Before joining Pennsylvania State University, Renfei Yan received a B.S. in Electronic Engineering from Peking University, and an M.S. in Biomedical Engineering from the Chinese University of Hong Kong. Her research interests span topics from biomedical signal processing to cost-effective health care delivery, health care intervention and management, demand forecasting, and economical modeling. Renfei Yan was a Scholar for the Center for Integrated Healthcare Delivery Systems in 2011. Her work on the performance evaluation of blood pressure measuring devices won the Student Research Award by Hong Kong Medical and Healthcare Device Industries Association. Based on this work, she drafted the IEEE Standard P1708 as the first author. She has been serving as the Student Representative for the administrative committee of IEEE Engineering in Medicine & Biology Society (EMBS) from 2010 to 2012. Currently, she is working at the Bristol-Myers Squibb as an Associate Manager in the department of Global Statistics.