

**CASES OF IMPROVEMENT TO PUBLIC HEALTH SYSTEMS
USING MATHEMATICAL MODELING**

A Thesis
Presented to
The Academic Faculty

by

Carlo Stefan Davila Payan

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2013

COPYRIGHT © 2013 BY CARLO STEFAN DAVILA PAYAN

**CASES OF IMPROVEMENT TO PUBLIC HEALTH SYSTEMS
USING MATHEMATICAL MODELING**

Approved by:

Dr. Julie Swann, Advisor
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Dr. David Goldsman
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Dr. Pinar Keskinocak
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Dr. Nicoleta Serban
School of Industrial and Systems
Engineering
Georgia Institute of Technology

Dr. Michael Washington
NCEZID/DPEI Health Economics and
Modeling Unit
*Centers for Disease Control and
Prevention*

Date Approved: November 4, 2013

To Norma, Sara, Rina, and Carlito; I love you all with all my heart.
To Mom, who could not wait any longer for me to finish this dissertation and went to
Heaven earlier this year.
To Dad, wishing you will stay around much longer.

ACKNOWLEDGEMENTS

I want to thank all of those who offered me their friendship and support during my years of study. Thank you Angie and Joel M., Judy and Mark H., Rinda and Tim L., Lory and David M., Amy and Erik V., Andrea and Tim K., Julie and Mark M., Elaine and Al L., Julie and Emil R.; you made us feel home in Atlanta. Thank you Westminster family, you will always be in our hearts.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	vii
LIST OF FIGURES	viii
LIST OF SYMBOLS AND ABBREVIATIONS	x
SUMMARY	xi
<u>CHAPTER</u>	
1 Introduction	1
2 Supply Chain and System Factors to Explain H1N1 State Vaccination Rates for Adults in US Emergency Response to Pandemic	4
Introduction	4
Methods	8
Results	14
Discussion	16
3 Supply Chain and System Factors to Explain H1N1 State Vaccination Rates for Children and High-Risk Adults in US Emergency Response to Pandemic	21
Introduction	21
Methods	24
Results	31
Discussion	35
4 Estimating Childhood High-BMI Baseline Prevalence in Small Geographic Areas	41
Introduction	41

Methods	46
Results	52
Discussion	59
Conclusion	61
5 Modeling Disruptions to Food Supply Chains in the U.S. caused by absenteeism	
During a Pandemic	63
Introduction	63
Literature Review	64
Methods	68
Preliminary Results	75
Discussion of preliminary results	86
APPENDIX A: Comprehensive List of Analyzed Variables Potentially Related to H1N1	
Vaccination Coverage	87
APPENDIX B: Variables Highly Correlated with Factors in the Adult Model	105
APPENDIX C: Variables Highly Correlated with Factors in the Children and High-Risk	
Adult Models	106
APPENDIX D: Comparison for obesity estimation using linear regression	108
APPENDIX E: List of Census Tables Used to Obtain Information of Small Geographic	
Areas	111
APPENDIX F: Model for Validation in Adult Population	114
REFERENCES	115
VITA	127

LIST OF TABLES

	Page
Table 2.1: List of variables in the final model for adults	15
Table 2.2: Regression results for predicting adult coverage	16
Table 3.1: List of variables appearing in children and adults models	32
Table 3.2: Regression results for predicting coverage for children	33
Table 3.3: Regression results for predicting coverage for high-risk adults	34
Table 4.1: List of variables appearing in children and adults models	48
Table 4.2: Adjusted coefficient estimates of the model for individual prediction (values of covariates were scaled to a [0,1] interval before regression) – Children 2-17 years old in Georgia.	53
Table 4.3: Examples of census tract prevalence estimates in Georgia, with 95% Confidence Intervals (C.I.)	54
Table 4.4: Examples of prevalence estimates for counties in Georgia, with 95% Confidence Intervals (C.I.)	55
Table 5.1: Number of nodes at each echelon of the milk supply chain	71
Table D.1: Linear regression model for adults	108

LIST OF FIGURES

	Page
Figure 2.1: Scheme of the supply chain of vaccine during the 2009 H1N1 pandemic response	5
Figure 2.2: H1N1 Vaccination coverage of adults by state as of January 2010	6
Figure 3.1: 2009 H1N1 vaccination coverage by state for 3 sub-populations	22
Figure 3.2: Example supply chain for H1N1 vaccine during the 2009-2010 pandemic vaccination campaign	23
Figure 3.3: Q-Q plots for the residuals of the children's and high-risk adult's models	35
Figure 4.1: Map of baseline overweight children prevalence estimation in Georgia by census tract	56
Figure 4.2: Map of overweight children population by census tract in Georgia	57
Figure 4.3: Map of children population prioritization by county in Georgia	58
Figure 5.1: Graphic representation of the milk supply chain network	71
Figure 5.2: Percentage of periods with unmet demand-base case	75
Figure 5.3: Variation on the percentage of periods with unmet demand-base case	76
Figure 5.4: Average days of inventory accumulated at each echelon-base case	77
Figure 5.5: Average days of inventory accumulated at echelons 1 – 6 (base case)	78
Figure 5.6: Average percentage of periods with inventory level of zero-base case	79
Figure 5.7: Average days of inventory at different levels of slack capacity and target inventory of 2 days	80
Figure 5.8: Average percentage of unmet demand per echelon at different levels of slack capacity and target inventory of 2 days	81
Figure 5.9: Average count of periods that inventory reaches a zero level at different levels of slack capacity and target inventory of 2 days	82
Figure 5.10: Average inventory level at a 10% slack capacity and variable target inventory	83

Figure 5.11: Average inventory level at a 10% slack capacity and variable target inventory	84
Figure 5.12: Average count of periods with zero inventory level at a 10% slack capacity and variable target inventory	85
Figure D.1: The 95% confidence interval for adult obesity county level estimates in Georgia, using linear regression models, compared with CDC data	109
Figure 5.12: The 95% confidence interval for adult obesity county level estimates in Georgia, using logistic regression models, compared with CDC data	110

LIST OF SYMBOLS AND ABBREVIATIONS

CDC	Centers for Disease Control and Prevention
HD	Health Department
MD	Medical Doctor
BMI	Body Mass Index
NHANES	National Health and Nutrition Examination Survey

SUMMARY

This work builds on the use of several Mathematical Modeling tools to develop approaches that address relevant, real and previously unanswered questions related to the improvement of Public Health Systems, in three particular instances.

First, this thesis analyzes the variation in state-level vaccination coverage during the emergency response to the 2009 H1N1 pandemic influenza outbreak in the United States. The analysis considers the overall adults population and two priority sub-populations: children and high-risk adults. We focus on quantifying the association between vaccination coverage and the supply chain and distribution system decisions, during the vaccine shortage period, while controlling for other commonly recognized factors such as previous vaccinations, socio-economic characteristics, health seeking behavior and health infrastructure. The variables analyzed are generally correlated, and the problem has a limited sample size with a much larger number of independent variables. The findings of this research have been published in *Vaccine* and presented to the Centers for Disease Control and Prevention.

Second, the research approaches the problem of estimating childhood obesity prevalence in small geographic areas in the U. S. Obesity is recognized as one of the major health problems in the country, and attending this condition in children is of major importance to deal with the sources of the overall problem. The ability to target interventions to the most affected children populations is necessary to achieve cost effective solutions. But local accurate obesity data is hard to obtain and missing for most of the small areas in the country. The research focuses on estimating prevalence of

obesity and overweight status in children in small geographical areas in the absence of surveillance and detailed sampling. Our modeling approach is built in two stages. The first one uses a logistic regression model that links individual characteristics to high-BMI status, and generates samples of the empirical distribution of its coefficients through bootstrap re-sampling. The second uses simulation to generate virtual population samples of the small areas, which are then combined with the logistic model samples to estimate prevalence. Confidence intervals are built through re-sampling. A very important feature of our approach is that all of its inputs are from publicly available data, which gives availability for the replication of the methodology to any health stakeholder in the US. The model estimates were validated by using separate models for adults and children in a state with available data. Estimates obtained from our modeling approach were used by a large healthcare provider to geographically target interventions for pediatric obesity.

Third, the thesis presents an introductory analysis of the possible effects of partial disruptions to critical supply chains due to absenteeism caused by a generalized flu-like illness in the US. For this analysis, we first construct a plausible national food supply chain for milk and then we simulate its disruption. To build the supply chain we used public information regarding production, consumption, and major milk processors and bottlers, and fitted it into a supply network through optimization. Then, to analyze the effects of flow disruptions of the supply chain, we built a simulation of the operation of the network and virtually generated absenteeism, mildly disrupting the supply chain flows by the proportional absences. We used information on potential absenteeism in work groups from an influenza simulator. Our initial analysis shows that absenteeism may create variations along the supply chain, similar to those described in the bullwhip

effect analysis literature, even in the absence of supply shortages and without variations in pricing or demand, for which we find no prior reference in the literature.

CHAPTER 1

INTRODUCTION

Traditionally, the methodologies used and developed by Industrial and Systems Engineering have been built and applied for industrial and commercial practices. It is not until recent years that problems in the public health arena are being formally incorporated as a new study subject by researchers in this engineering area. Public health systems in the United States have a vast number of operations and preparations for possible scenarios that will admit much collaboration of research with industrial and other engineering areas for years to come. This thesis presents three instances in which mathematical tools commonly used in Industrial and Systems Engineering analysis have been updated, enhanced, and used for the improvement of public health systems.

The first public health systems problem analyzed is aimed to understanding the causes of the differences in the state-level vaccination coverage for the Novel H1N1 Influenza pandemic that affected the US at the end of 2009, and beginning of 2010. This study performed an analysis of the relationship between the state-level percentage of vaccination in 3 sub-populations (overall adults, high-risk adults and children), and more than 180 potential covariates, including many indicators of logistic performance during the distribution of the vaccine and supply chain decisions made by each of the state's governments. Such an extensive approach became a comprehensive explanation that relates vaccination rates with supply chain decisions, while controlling for factors commonly related to vaccination. The results were presented to an important public

health institution, the Centers for Disease Control and Prevention (CDC) and published by Vaccine journal.

The findings from the analysis of this historic vaccination effort added up to the lessons learned in the aftermath of the massive campaign. Some of the results confirmed CDC's educated intuition. But some other findings became apparent only after they were exhibited by the study. Some outcomes challenged common decision policies, and pointed the way to better practices. The approach presented in the analysis of this first problem offers a straight forward way to deal with the technical complications of building linear regressions when the number of independent variables is much larger than the number of observations, and the independent variables are highly correlated. Chapter 2 will present the study for overall adults, while chapter 3 will present the analysis on two priority subpopulations: children and high-risk adults.

The second analysis is related to one of the most relevant public health systems problems in our days: Childhood Obesity. Targeting interventions to reduce childhood overweight/obesity requires finding those sub-state-level areas where the percentage of obese children is higher. But finding prevalence estimators for such small regions is not trivial. In fact, there are no previous publications proposing methodologies to estimate prevalence of childhood obesity in US small areas for children under 12 years old, other than applying direct sampling, physically measuring children, which can be considerably expensive. Different from adults, telephonic surveys have proven to be ineffective to predict children's obesity status. This thesis presents a modeling strategy that can be used to estimate childhood overweight/obesity at a zip code or census tract level using publicly available data.

This estimation approach was used in Georgia. The results were found to coincide with the state-level point estimation for children 10–17 years old, and the approach was validated by comparing outcomes for adults at a county level with those of available surveys for Georgia, and outcomes for school age children in Arkansas counties to the information obtained in that state by measuring children in schools. This work was used by a healthcare provider to target interventions in the state of Georgia, and may provide relevant insights for improving more public health systems in the future. Chapter 4 contains the analysis of this problem.

The third and final problem, presented in Chapter 5, analyzes to the possible effects of absenteeism due to a generalized flu-like illness to critical supply chains in the United States. The first approach considered modeling some food items' supply chains, without changes to supplied or demanded quantities. A first set of results shows that absenteeism caused by a flu pandemic can cause similar effects to those caused on supply chains by changes in supply or demand. We show that tactical responses to shortages through the application of inventory and ordering policies and slack capacities across the network cause a characteristic distribution of the inventory and service levels across the different supply echelons.

Industrial and Systems Engineering encompasses an extensive variety of mathematical and systemic approaches that can be developed to bring improvement to public health systems. It is relevant and necessary that research in industrial engineering, related to the different aspects of health related systems, continues its development. Hopefully, the present work will serve to foster future related research.

CHAPTER 2

**SUPPLY CHAIN AND SYSTEM FACTORS TO EXPLAIN H1N1
STATE VACCINATION RATES FOR ADULTS IN US EMERGENCY
RESPONSE TO PANDEMIC**

Introduction

The novel H1N1 influenza virus was detected in the United States in April 2009. Worldwide, a pandemic was declared, and a national public health emergency was announced in the United States. In the US, plans were made for a national vaccination campaign to be rolled out in Fall 2009, when the pandemic H1N1 vaccine would be available. The campaign was implemented as a public-private partnership, with federal purchase of the vaccine. The Centers for Disease Control and Prevention (CDC) allocated vaccine pro rata to states by total population as the vaccine became available. States determined how vaccine would be allocated in their jurisdiction and either retained control of vaccine allocation to individual providers at the central level or delegated fully or partially to local jurisdictions. States or local jurisdictions invited providers to participate in the program and vaccine was shipped to designated providers through a centralized distribution process supervised by the CDC that built on an existing contract for management and distribution of vaccines in the Vaccine for Children (VFC) program. Fig. 2.1 shows a basic scheme of the supply chain for H1N1 vaccine from manufacturer to provider.

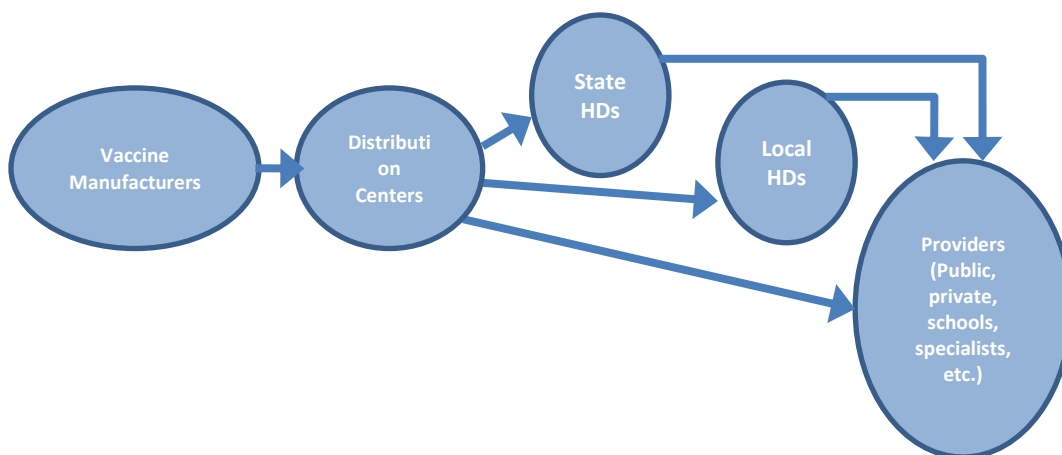


Figure 2.1: Scheme of the supply chain of vaccine during the 2009 H1N1 pandemic response.

State decisions about where to direct vaccine were guided by recommendations of the CDC’s Advisory Committee on Immunization Practices (ACIP) (1), which recommended that the vaccine be initially directed to: pregnant women, persons who live with or provide care for infants aged <6 months, health-care and emergency medical services personnel who have direct contact with patients or infectious material, all people 6 months to 24 years of age, and persons aged 25 through 64 years with certain health conditions (“high-risk”). The recommendations also provided further specification of priority groups in the event of vaccine shortage and stated that decisions to broaden availability of vaccine should be made at the local level.

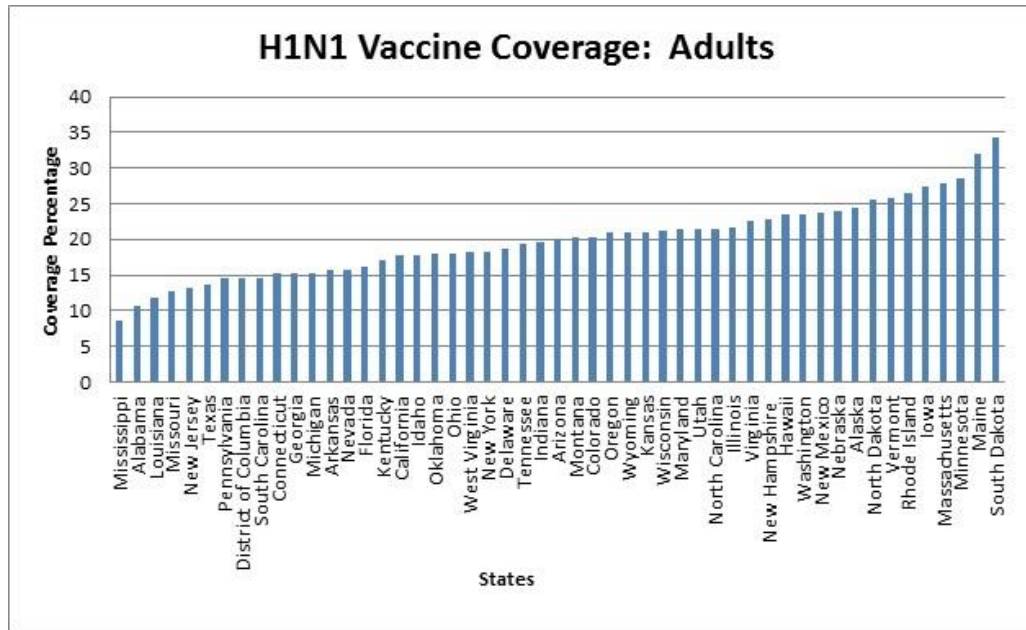


Figure 2.2: H1N1 Vaccination coverage of adults by state as of January 2010.

Overall, more than 120 million doses of vaccine were distributed to over 70 thousand locations by April 2010 (2-4) and 80.8 million people reported having been vaccinated (5). The vaccine supply was insufficient to meet demand initially, and became more plentiful after Thanksgiving, a time when demand for influenza vaccination traditionally slows. Despite the pro rata allocation of H1N1 vaccine (6), state level vaccine coverage rates indicate that there were great differences in coverage across the states even when vaccine was in short supply. By the end of January 2010 (7), the coverage of adults ranged from 8.7% to 34.4% (Fig. 2.2).

States varied in their approaches to implementing their H1N1 vaccination programs in an unprecedented situation. While the literature addressed factors related to uptake of seasonal influenza vaccine at the individual level (8, 9), states and regions used their best judgment and knowledge of their jurisdictions to guide their decisions on

distribution and system design, given the lack of scientific evidence in that area. The purpose of this study was to determine supply chain and system factors associated with H1N1 coverage rates at the state-wide level for adults in order to inform future events of this nature.

We hypothesized that characteristics of the vaccine supply chain in each state and decisions around targeting vaccine could predict uptake. One classic supply chain study, for example, has demonstrated that a product stocked in a large number of locations increases the probability that a particular location will be stocked out, and may also reduce the distance traveled by the final consumer (10). Some of these characteristics of the state vaccine supply included the number of locations where vaccine was available, prioritization of the ACIP-recommended target groups, the type of providers to whom vaccine was directed, and the lead-time between vaccine allocation and availability in a state, which largely reflects differences in states' ordering processes.

Because other factors affect uptake, as evidenced by state-to-state variation in seasonal influenza coverage and individual-level studies (11-14), underlying population differences such as demographic characteristics, utilization of preventive health services, and healthcare infrastructure were also examined. It is relevant to mention that individual-level studies differ from those with a regional or ecological view. Others have used this ecological approach in the analysis of other health-related problems such as water fluoridation and tooth decay (15, 16). Data from the centralized distribution system on vaccine shipments from October 5, 2009 through December 9, 2009 were made available for analysis, thus allowing us to focus the analysis on the period during which vaccine was in short supply.

Methods

Data

The outcome measure is state estimates of vaccination coverage, as calculated by the CDC (7). Participants 18 and over on BRFSS and NHFS surveys were asked if they had received an H1N1 vaccine during October 2009-January 2010.

Population and state characteristics

From the Census, we identified population (17) and socioeconomic characteristics including population size and density, and its composition by age groups, education, race/ethnicity, income and poverty, births, and family composition (18). We took measures of income inequalities from a study conducted by the University of Connecticut (19), and measures of segregation and disparities from work by the Kaiser Family foundation (20).

We obtained data on state characteristics from several sources. We extracted the geographical area, the number of counties, and the federal government expenditure per capita from the Census, and the number of cars in the population from an industry trade report (21). We determined whether states had “home rule” , defined as whether states were characterized by state control, local control, or by inference, mixed control, from the 2008 National Profile of Local Health Departments (22).

We estimated the total number of healthcare practitioners and the number of active physicians per thousand population (PTP), respectively, from the Bureau of Labor and Statistics (23) and the American Medical Association 2006 (24). We obtained data on the size of the medically underserved population and the percentage of the population who have not visited a doctor in the last year because of cost from State Health Facts (25). From the CDC’s Behavioral Risk Factor Surveillance System (BRFSS) (26), we

extracted the state specific population percentages associated with various health conditions.

We used several sources to quantify health-seeking behaviors and use of preventive services. We obtained state-specific influenza vaccination rates for previous seasons for different population groups from CDC's BRFSS (27). We also considered the percent of adults with a dental visit in the past year (28), the percent of women who had a Pap smear in the past 3 years (25), and the percentage of adults aged 65+ who have ever had a pneumococcal vaccine (25).

State-specific vaccination program and surveillance information

We obtained information on the emergency funding provided to states for the H1N1 pandemic from CDC reports (29): Total (Public Health Emergency Response (PHER) funds per capita in 2009; unobligated percentage for PHER I and II (awarded in July and August for assessment and planning, respectively) to determine the amount spent; and total or unobligated Phase III Public Health Emergency Response (PHER) funds per capita (30) (awarded in September 2009 for implementation of campaign).

Reports from the Outpatient Influenza-like Illness Network (ILINet) (31) obtained from the CDC, provided weekly values for the proportion of outpatient visits for influenza-like illness (ILI) at participating providers, by state, from which we calculated the week of peak prevalence in Fall 2009, and the length of time with major ILI activity by several different measures: Peak week for number of ILI cases after week 30; peak number of cases after week 30, per thousand population; percentage of weeks with % ILI above 2.3, after week 30; percentage of weeks with number of cases above 1/3 of the peak; number of weeks from peak until ILI dropped below 1/3 (or 1/2) of the peak.

We extracted information on state processes and decisions from a survey of immunization program managers conducted by the University of Michigan to provide

CDC with situational awareness during the H1N1 campaign (32). Findings from the survey were used to categorize states according to who approved providers (state or local public health), whether vaccine allocation to private providers was made at the state or local level, early allocation strategies during the shortage period (e.g., pro-rata, some per county, focused on target populations), whether or not vaccination had expanded beyond the ACIP target groups by Dec 4 (similarly, by Dec 18), use of retail vaccinators as of December 18 (widely, minimal, later or not planned), and whether or not (states, or local health departments, or third parties) distributed vaccine directly to providers (i.e., vaccine that was not shipped directly from the centralized distribution system).

We obtained information on the amount of vaccine allocated to each state over time, the maximum number of provider sites to which each state could have vaccine shipped through the centralized distribution system (“ship-to” sites) (3), the number of providers/practices that had entered into an agreement by 12/16/09 with their state health department to provide H1N1 vaccine (33), distribution center assignment (which of four) (3), the number of vaccine doses received in each state through the federal pharmacy vaccination initiative (34) in late 2009, and self-reported data from states on doses distributed to or administered in public settings (35), all from internal CDC reports.

Information on the date, address, and number of doses shipped to each location, from the beginning of the campaign through December 9, 2009 (which covers the major shortage period) was obtained from the centralized distribution shipping records (2). This allowed us to calculate the number of unique sites to which vaccine was shipped (ship-to-sites), the number of shipments PTP, and the average number of shipments per site. We also calculated the lead-time from allocation to shipment (i.e., the average number of days between when a state received an allocation and ordered the vaccine, plus the average number of days doses spent between order placement and shipment), and the ordering frequency per week. Shipments during this time period were sent overnight to

their destination (regardless of distance), to arrive when receiving locations within the state were open. Lastly, we calculated the variation in doses PTP across counties within a state, by averaging the absolute value of the difference between the overall state doses PTP minus the doses PTP in each county.

We categorized shipments by the type of provider (e.g., pediatrics or children's clinic, one of several kinds of specialists, local health departments) through a series of targeted queries based on name. From this derived data, we calculated proportion of shipments or doses PTP to providers focused on children, primary care, county health departments, unclassified medical doctors, internists, specialists, long-term care, veterans, urgent care, hospitals, clinics, pharmacies, jails, military, government, universities, and nursing homes; we also combined these in several subgroupings driven by like characteristics that might explain differences in coverage: general internists and specialists combined (internists and specialists can be grouped because both serve adults; however, while internists may provide primary care, adults may be less likely to visit internists or specialists during a short campaign); targeted access (doses sent to long term care, internists, specialists, nursing homes, and children); open sources and children (doses sent to counties, primary care, and children); all locations other than children or county; general access locations (primary care, MDs that could not be classified by specialization, counties, hospitals, urgent care, clinics, or pharmacies); and restricted access institutions (universities, military, jails, government, companies and veterans). Information was adequate to categorize more than 75% of the overall shipments. Appendix A provides a comprehensive list with variables considered, and their sources.

Analysis

The variable selection process continues to be a recurrent and difficult problem to deal with (36), and regression requires making decisions on variable selection approaches and other modeling perspectives (37). We certainly avoided trying to consider every combination of every set of variables and pick the model with the highest R-squared (or best Akaike information criterion (AIC) value), which would pose a large combinatorial problem. Although we considered some automatic variable selection algorithms during the exploratory phase, we found them not effective for dealing with our highly correlated variable set. Thus overall, we used stepwise approaches, evaluating at each step both the explanatory power of additional variables (and their significance in the model) along with the correlations among variables.

Statistical theory indicates that it is relevant to cluster highly correlated variables and select at most one representative variable for each of these clusters to test into the regression model (see for example Doctoral Thesis work by S. Park (38)). Also, it has proven useful for the construction of models to select the regression variables while controlling for interrelationships among them (39). We followed these principles, and enforced them in our implementation in three ways: 1) by selecting initial sets of variables that were not highly correlated among them, but correlated with H1N1 coverage, 2) by restricting the variables that had a high correlation with any other independent variable in the model from entering into the model, and 3) by allowing swaps of highly correlated variables. Below we give a more detailed description of our process for obtaining a final model:

1. We quantified the group of variables that were the most highly correlated (either positively or negatively) with adult coverage. Examples of these include: 1) socio-economic variables such as percentage of population who is black (variable c4) and

percent high school graduate or higher (c10); 2) health mortality or morbidity such as cardiovascular deaths (ah10), mortality with end-stage renal disease (ah43), adults with diabetes (ah2), good or better health reported (ah4); 3) use of health services such as previous flu vaccinations (e.g., sf2, sf3, and sf7) and percentage reporting not seeing a doctor because of cost (ah25); 4) H1N1 epidemiology, the percentage of weeks with % ILI above 2.3 after week 30 (pw3); 5) H1N1 campaign, the average number of days between allocation and shipment of vaccine (o7).

2. Many of these variables (e.g., socio-economic or mortality) had high correlations with each other or with many other potential variables. Some variables, such as percent of black population or education attainment had many collinearities, making it difficult to develop strong models from them. Therefore, variables with many collinearities were not chosen as starting points for the models.

3. We chose one small initial set of variables at a time to begin model development (could be a single variable), where we selected from among categories or variables that were not too highly correlated with many other potential dependent variables, but still had a significant correlation with the independent H1N1 coverage variable. Examples of good starting variables were previous seasonal flu coverage variables. For each startup model, we added one variable at a time (stepwise forward selection, using the `add1` function in R), where we chose new variables with the highest potential to benefit the model from a set of potential variables while preventing the inclusion of any variable that was highly correlated with those already in the developing model. With each new model, we checked for potential variables to remove by examining information such as p-values and measures of potential benefit to the model (reduction of AIC, using the `drop1` function in R). We also considered swapping variables in a model by highly correlated potential variables if the latter consistently pointed to be of high potential benefit to the

model. We iterated many times with steps forward and backward until arriving at a model without obvious improvements.

4. We began the model development process again, by starting with a new set of initial variables, iterated, until arriving at another model without obvious improvements. Models with higher adjusted R-squared were kept and models with lower adjusted R-squared were discarded.

5. Although it would be virtually impossible to analyze all possible combinations of the variables, we observed after a few repetitions that some variables would consistently find their way into the models, and actually appeared in our final model, sometimes even without having some of the highest individual correlations to H1N1 coverage.

Results

Seven variables including lead-time from allocation to ordering and shipment, the maximum number of ship-to sites per thousand population, past seasonal influenza coverage for non-high risk adults age 18 -- 49, percentage of doses categorized as sent to internists and specialists, percentage of women 18 and older with a pap smear in the last three years, percentage of weeks with ILI above 2.3 after week 30, and the percentage of residents of Hispanic or Latino origin were significant for predicting vaccination coverage in adults (Table 2.1). The best model found explained the variation in state-specific adult vaccination coverage with an adjusted R-squared of 0.76 and a p-value close to 0 (Table 2.2).

For supply decisions, a long lead-time was associated with lower coverage, and the associated coefficient has a relatively large magnitude. Additional analysis of lead-time indicated that a state's relative lag tended to be consistent throughout the months considered. We also found that lead-time is correlated with some variables related to shipment choice (e.g., positively with use of third parties for distribution, and negatively

with shipments per ship-to site). The vaccine allocated to internists and specialists as a percentage of the total shipped was negatively associated with coverage, and having a large number of maximum ship-to sites was positively associated with coverage.

Table 2.1: List of variables in the final model, including the dependent variable at the top. Table shows the variable's name, description (with reference for the data), average value, standard deviation, and maximum and minimum values.

Variable	Description	Reference	Average	Std. Dev.	Maximum	Minimum
Dependent	Coverage on persons aged \geq 18 yrs.	MMWR(7)	19.9	5.3	34.4	8.7
Indep1	Percent of Women Age 18 and Older Who Report Having Had a Pap Smear Within the Last Three Years, 2008	State Health Facts(25)	82.7	2.9	88.9	74.1
Indep2	Resident population: Hispanic or Latino Origin, percent (July 1 2009 - estimate)	Census(17)	9.8	9.6	44.9	1.1
Indep3	Average days from allocation to shipment of vaccine	CDC Shipments Report (Calculation) (2)	6.3	2.3	12.5	2.1
Indep4	Percentage of weeks with % ILI above 2.3, after week 30	Report CDC(31)	42	24.4	97.4	10.3
Indep5	Seasonal influenza Coverage for non-high risk adults 18 -- 49 yrs. on the 2007-2008 season	CDC Influenza Vaccination Coverage(27)	22.6	5.1	37.8	11.9
Indep6	Maximum number of ship-to sites per state per thousand population	Report CDC (Calculation) (3)	0.5	0.1	0.74	0.00*
Indep7	Percentage of doses categorized as sent to internists or specialists**	Report CDC (Calculation) (2)	1.17	1.08	6.07	0*

Table 2.2: Regression results for predicting the state level vaccination coverage for the adult population. The table contains the variable name, short description, point estimation of the variable's coefficient, coefficient's standard error, coefficient's t-value, and results of the significance test.

Variable	Short Description	Estimate	Std. Error	t value	Pr(> t)	
	(Intercept)	2.66E-16	0.06807	<0.001	1.00E+00	
Indep2	% Hispanic	0.378	0.07953	4.753	<0.001	***
Indep5	Past influenza Coverage	0.3599	0.07928	4.54	<0.001	***
Indep1	% Women w/ Pap	0.3002	0.07653	3.923	<0.001	***
Indep6	Max # Sites	0.1807	0.07061	2.558	0.01412	*
Indep7	% to Specialists	-0.295	0.07788	-3.788	<0.001	***
Indep4	% weeks ILI high	-0.4366	0.07362	-5.931	<0.001	***
Indep3	Lead-time	-0.4419	0.07401	-5.97	<0.001	***

Significance codes: 0 < '***' < 0.001 < '**' < 0.01 < '*' < 0.05
Adjusted R-squared: 0.7637, Reg. p-value: 6.035E-13

Vaccination coverage was positively associated with past influenza vaccination coverage; while we found a strong association, there were several other effects that were also large in magnitude. Coverage was also positively associated with the percentage of women with a pap smear, and the percent of the population that is Hispanic. A long duration of ILI severity peaks (defined by the percentage of weeks in the Fall with percent ILI more than 2.3) was negatively associated with coverage. To provide more information on our modeling, a supplementary table in Appendix B presents examples of other variables highly correlated with those factors in our final model.

Discussion

In an effort to identify lessons learned for a future pandemic vaccination event, we sought to identify factors related to vaccination program decisions and processes that may have facilitated or hindered vaccine uptake. Program factors that were associated with vaccine uptake included the lead-time between allocation and ordering and shipping,

and the type of providers receiving vaccine. Factors not related to program decisions such as health-seeking behaviors and population characteristics also contributed to predicting state-to-state variation, as would be expected given baseline variation in previous influenza vaccination coverage(27) and other findings (40-42).

Lead-time from allocation to ordering and shipment was negatively associated with vaccination coverage. Steps in the ordering process varied by state and could include requesting specific orders from providers (in advance of allocation or after receiving an allocation), decisions on where to distribute vaccine, and notification of decisions. States also determined the frequency of ordering, the day(s) of the week to order, the number of providers participating or receiving vaccine, and the overall process to follow, all of which could affect the lead-time. Because of the initial focus on ACIP-defined target groups, in many states adults without high risk conditions were not eligible for vaccination until demand for vaccine had already begun to wane. Delays in allocated vaccine being made available to the population could have resulted in less vaccination. On the other hand, lags in ordering could be a consequence of decreasing demand, and thus be a result of lower vaccination rates rather than a cause. The tendency for lags in ordering to be consistent for a given state throughout the time period studied, suggests the lead-time resulted from the ordering process.

We also found a relationship with the type of providers or locations to which vaccine was directed. For adults, vaccine sent to providers with specialized services or patient base was associated with lower coverage. This could be because not all adults visit internists or specialists frequently enough to be vaccinated in this time period; it could also be that those providers had less focus traditionally on vaccinating so patients

looked elsewhere for vaccine. Overall, only a small proportion of vaccine was sent to internists and specialists.

One variable may be more a measure of health infrastructure than the supply chain system itself. In particular, the maximum number of sites to which vaccine could be directly shipped through the centralized distribution system) was positively associated with vaccination coverage. (In contrast, another variable measured the actual ship-to sites registered or used within a state.) The maximum number of ship-to-sites allowed for each state was based on a formula that included the population size as well as the number of existing VFC providers. A high number of VFC sites per capita could be a reflection of a more robust infrastructure for providing vaccine.

State factors unrelated to supply chain decisions about H1N1 vaccine were also related to coverage, specifically included usage of health services. Others have found that for an individual, past influenza vaccination is a strong predictor of annual influenza vaccination (8, 13): a relationship that may reflect both differences in infrastructure and differences in attitudes. The finding in this paper demonstrates that pandemic influenza vaccination also is associated with uptake of seasonal vaccine. The association between coverage rates and rates of receipt of Pap smear may be a reflection of utilization of preventive care, although no further analysis could be carried out to determine if this effect was present only among women.

Some characteristics of the epidemic may have also influenced coverage. For states where the epidemic lasted longer, coverage was lower. This could be because vaccine was made available to non-high risk adults later in the season, and persons may have reasoned that they had likely been exposed to the disease already and did not need

vaccination. Conversely, the positive association between coverage and the percentage of Hispanics may reflect higher vaccination rates in communities with greater perceived risk (43) due to the virus emerging from Mexico. In general, Hispanic populations did not have a higher coverage than the overall average. (44)

This study had several limitations. First, cross sectional studies and regressions are useful for identifying associations, but they have a number of intrinsic limitations, for example, we cannot determine causality, and for complex cases like the one analyzed other good regression models may also exist for the same set of variables. Secondly, the ecological approach followed does not point to individual characteristics of the population but to state-level conditions, and does not analyze potential variations within states. Third, the data from the centralized distribution system covers shipments through December 9, 2009, and the outcome measure is vaccination coverage as of the end of January 2010. The gap may not be as large as it seems, since coverage for adults increased from 17.3% (adults ≥ 19 (45)) at the end of December 2009 to around 18.2% (adults ≥ 18 , derived from state-specific rates (7) and adult populations (17)) at the end of January 2010. Additionally, the number of people vaccinated by the end of January (74M) is approximately the same as the total vaccines shipped by December 9 (72M) though this comparison does not take into account receipt of second doses by children. Fourth, the vaccine shipment data represented shipment location, which is not necessarily the same as the final place of administration of vaccine (e.g., vaccine may have been distributed from a third party distributors or local health department to providers). As a result, the number of locations of administration may be underestimated, or the provider type may be misclassified. Fifth, some shipping data were missing or potentially

inaccurate. Provider type could not be determined for 25% of shipments, the information on state and local decisions and processes was not always complete, and databases could have errors. Finally, the number of dependent variable observations is fairly small (51), and many factors may potentially be associated with H1N1 coverage.

The distribution and administration of the H1N1 vaccine was a test of the health emergency response systems, and it is an opportunity to identify specific approaches that may result in higher vaccine uptake in a future event of this nature. Several of the findings warrant further consideration. The findings suggest that continued efforts to increase uptake of influenza vaccination may result in increased uptake in an emergency response. The negative association between order lags and coverage is an important aspect of the supply chain and distribution. It is possible that time lags are a function of the system design or processes, which would suggest monitoring and/or designing the system for fast response within the states in an emergency is needed. There can be many decisions made at the state level that can affect lead-time including ordering frequency, number of delivery locations, on which days orders were placed, use of third parties, etc. Further study would be useful in this area. Our results on type of location to which vaccine was directed may provide some guidance on increasing coverage, e.g., in a campaign with limited resources and time pressures, sending to general access or public locations may be beneficial. As more adult and specialty providers, including pharmacies, take on the role as vaccinators, this strategy may change. This, too, remains an area where additional analysis is useful, such as collecting information on shipments by type of provider, examining the small number of states where registry information records the location of vaccine administration, or additional analysis on where vaccination occurred for different target groups.

CHAPTER 3

**SUPPLY CHAIN AND SYSTEM FACTORS TO EXPLAIN H1N1
STATE VACCINATION RATES FOR CHILDREN AND HIGH-RISK
ADULTS IN US EMERGENCY RESPONSE TO PANDEMIC**

Introduction

A national vaccination campaign was rolled out in the fall of 2009 in response to the H1N1 influenza pandemic. Initially, the vaccine was in short supply, in some areas until early December. The vaccine was purchased by the federal government and allocated to states as it became available, in proportion to population size. The flow of doses from the manufacturers to the national distribution centers and then to final points of distribution built on an existing contract for management and distribution of vaccines in the Vaccine for Children (VFC) program. Depending on their internal structures, states or local authorities decided how to distribute vaccine within their jurisdiction.

CDC's Advisory Committee on Immunization Practices (ACIP) issued recommendations for the use of the vaccine (46). The initial target groups were: pregnant women, household contacts or caregivers for infants aged <6 months (e.g., parents, siblings, and daycare providers), personnel who have direct contact with patients or infectious material at health-care and emergency medical services, all people between 6 months and 24 years of age, and persons 25 through 64 years old with health conditions associated with higher risks of complications if infected (to whom we will refer as "high-risk adults") (46). The recommendations further specified priority groups in the event of a

vaccine shortage, giving priority to the first three of the previous groups, and in addition children aged 6 months—4 years, and children and adolescents aged 5—18 years who have a medical condition that could cause them influenza-related complications. Finally, the ACIP recommendations stated that decisions about opening vaccination up beyond the target groups should be made at the local level.

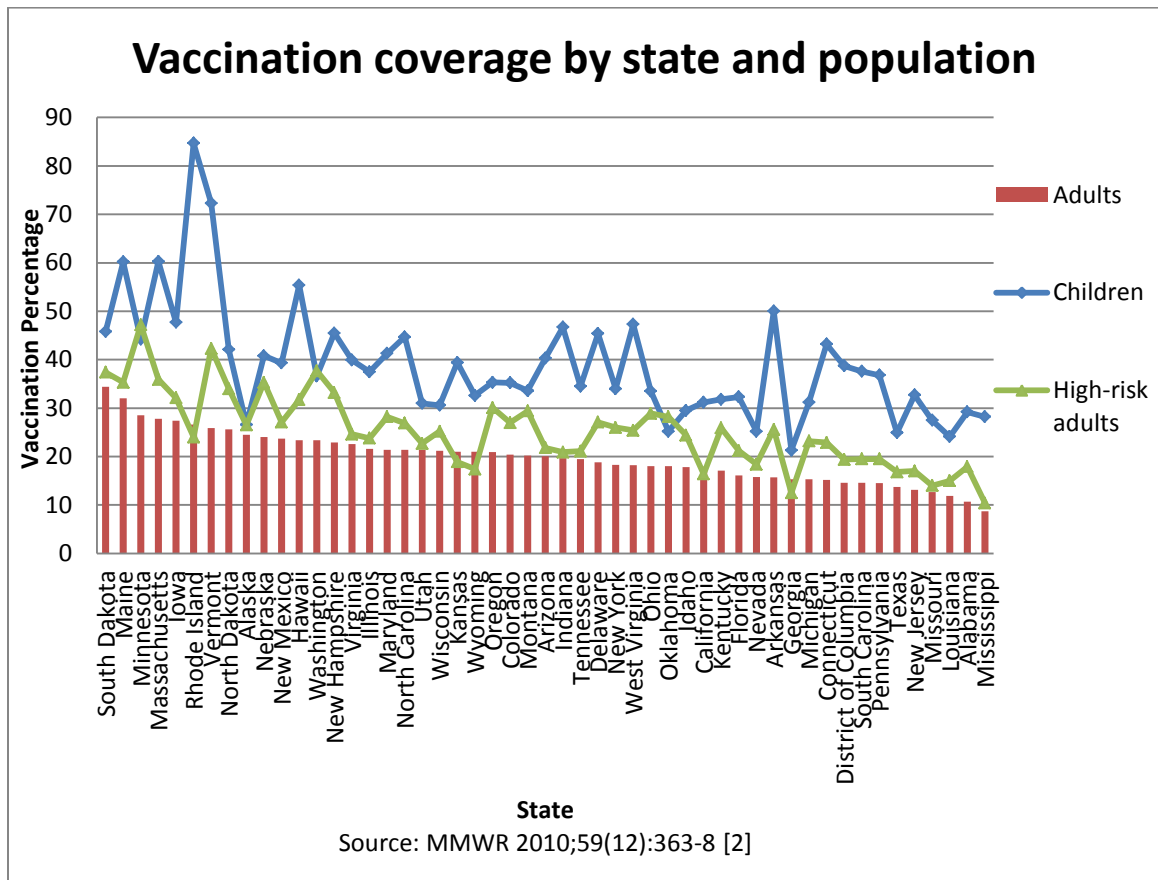


Figure 3.1: 2009 pandemic H1N1 vaccination coverage by state for 3 different populations (7), sorted by decreasing order of adults coverage, October 2009-January 2010.

Despite the pro-rata allocation of vaccine to the states, by the end of January 2010 (7) state-level vaccination coverage varied markedly across states, with rates for children

aged 6 months to 17 years ranging from 21.3% to 84.7%, and for high-risk adults from 10.4% to 47.2%. This variation suggests that implementation strategies (e.g. location of vaccination or types of providers receiving vaccine) may have affected state-level vaccination rates achieved and that specific distribution strategies may be associated with reaching specific groups. Figure 3.1 summarizes coverage outcomes (7) for children and high-risk adults compared to overall adults (18 and up, including those with high-risk conditions). Coverage rates were higher for more than one group in some states, pointing to the potential contribution of state systems, processes, or underlying characteristics to coverage achieved.

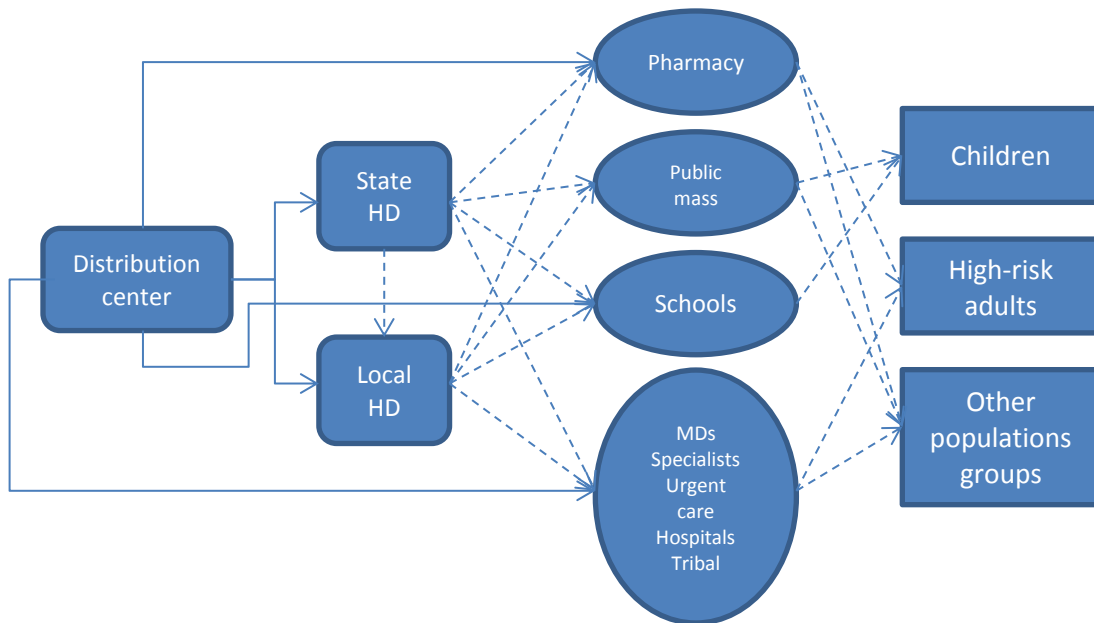


Figure 3.2: Example supply chain for H1N1 vaccine during the 2009-2010 pandemic vaccination campaign (Abbreviations: HD-Health Department; MD-Medical Doctor)

In a previous study, we found that certain supply chain and system factors were associated with state-level coverage of overall adults (47). The purpose of this study was to extend that analysis and focus on factors associated with coverage of children and high-risk adults, two of the initial target groups for vaccination. Some of the characteristics of the state's health supply chain that we expected to relate with coverage of children and high-risk adults were the number of locations where vaccine was available, type of providers that received doses, focus on school vaccination, timing of opening of vaccine distribution to non-priority groups, use of third parties for transfer and redistribution of vaccine, and use of retail and pharmacy for vaccination. Figure 3.2 represents an example of the construction of the supply chain of vaccine. We considered health infrastructure characteristics for the states, and data about vaccine shipments and distribution strategies during the primary shortage period. To account for other factors that may affect vaccination coverage (40, 42, 48-51), we included factors pertaining to the underlying characteristics of the state's population such as demographics and utilization of preventive health services.

Methods

Design

We used linear regression models to perform an ecological analysis on the relationship between state-level 2009 H1N1 vaccination rates in children 6 months to 17 years of age, and high-risk adults 25-64 years old, and variables describing 1) the state's supply chain and process design for the vaccination campaign, and 2) general population and health characteristics of the states.

Data

We separately analyzed two outcomes, both related to the state-specific 2009 H1N1 vaccination coverage: (i) the estimation of children's vaccination rate as a percentage (0% – 100%) of the population, and (ii) the estimation for the percentage of high-risk adults vaccinated, both of them calculated by the CDC(7, 52).

Population and state characteristics

The data sources for the analysis were varied including census(17, 18), income inequalities(53), measures of segregation and disparities(54), industry trade reports on number of cars(21), the 2008 National Profile of Local Health Departments(22), the Bureau of Labor and Statistics(23), the American Medical Association 2006(24), State Health Facts(25), CDC's Behavior Risk Factor Surveillance System (BRFSS)(26), and CDC estimates on influenza coverage for previous seasons(55)). The details on this data (and all others) are explained Appendix A.

For the analysis of children, we additionally considered several variables from the National Survey of Children's Health 2007 (56) that describe the children's general health condition, the prevalence of chronic health conditions among them, their private or public health insurance coverage, if they have preventive visits to the doctor in the past 12 months, and if their home meets the medical home criteria.

To calculate the state-specific 2009 H1N1 estimation for the percentage of high-risk adults vaccinated, participants on BRFSS and NHFS surveys administered during November 2009-February 2010 were asked if they (or their children) had received an H1N1 vaccine during October 2009-January 2010. High-risk conditions in the surveys include anemia (including sickle cell), asthma, diabetes, coronary heart disease, kidney problems, lung problems other than asthma, myocardial infarction, and a weakened immune system caused by illness or medicines (7, 52).

State-specific vaccination program and surveillance information

The analysis included information on emergency response funds provided to states(29, 30); reports from the Outpatient Influenza-like Illness Network (ILINet)(31); information on the amount of vaccine allocated to each state over time; detailed vaccine shipping information including date, address, and number of doses shipped to each location, from the beginning of the campaign through December 9, 2009(2) (which covers the major shortage period); the maximum number of provider sites to which vaccine could be shipped through the centralized distribution system; the number of vaccine doses received in each state through the federal pharmacy vaccination initiative(34, 57) in late 2009; and self-reported data from states on doses distributed to or administered in public settings(35).

Information on state processes and decisions from surveys during the campaign(32) contained several variables that we considered including: the percentage of VFC providers who participated in the H1N1 campaign; whether or not vaccination had expanded beyond the ACIP target groups by December 4, 2009 (similarly, by December 18); whether school clinics had been held by October 27 (or doses were being held for, or waiting to hold clinics at schools) and whether school vaccination was a main focus; and if 3rd party distribution was used to transfer or redistribute to small providers. When analyzing data from the Simultaneous Tracking Project led by Clark (32), we did not include in the analysis those variables that had a considerable number of missing data points, given that missingness can deeply affect the adequacy of our models.

From the detailed shipping information we calculated the average number of shipments per location (the total number of shipments divided by the total number of ship-to-sites per state). Performing targeted queries, we also categorized shipments by

type of provider, showing that vaccine went to pediatricians or children's clinics, primary care facilities, county health departments, internists, specialists (not including specialists for children or women), long-term care, veterans, urgent care, hospitals, clinics, pharmacies, and other venues that could be more closely related to the vaccination of children and high-risk adults. We also combined some of these categories in subgroupings to see which had a greater impact on these populations. For example: the targeted access group included doses sent to long term care facilities, internists, specialists, nursing homes, and children; the general access group include doses sent to primary care, MDs, counties, hospitals, urgent care, clinics, or pharmacies. Information was adequate to categorize more than 75% of the overall shipments.

Analysis

We constructed separate models for children (6 months to 17 years) and high-risk adults (25 to 64 year olds with a chronic condition) because we expected factors affecting coverage to differ across groups, and to differ from factors associated with vaccination rates in overall adults (18 and up, including those with high-risk conditions (47)).

The primary technique used for modeling was multivariate linear regression (ordinary least squares). We used a logarithmic transformation of the vaccination rate for children, to better approximate normality. We calculated simple descriptive statistics for all the analyzed outcomes and factors (means, standard deviations, and proportions). Outliers were not removed for the analysis. Data was linearly scaled to values in [0, 1] before performing regressions.

We selected a number of potential initial predictors for each of the dependent variables based on their correlation with the outcomes. From these initial models we

developed models by stepwise addition, elimination, or by interchange of factors. At each stage, we chose variables to include or remove based on their statistical significance and their potential to explain variability, while we examined correlations to avoid high collinearities in the model. Models were evaluated on adjusted R-square values and the F-statistic, with individual variables significant at $p\text{-value} < 0.05$. The regressions were performed with R statistical software package version 2.11.1(58). Some descriptive statistics were calculated in Microsoft Excel versions 11 and 12.

Coverage of 2009 H1N1 vaccination was measured as the percentage (0%-100%) of population who reported having been vaccinated. The outcome for high-risk adults was used as the dependent variable for the linear regression model built for that subgroup. The outcome for children was first transformed by calculating the natural logarithm of the coverage percentages. For example, a 30% coverage for the high-risk population, would have been assigned the value 30 for modeling, while a 30% coverage in the children model was transformed to $\ln(30) = 3.401$ for modeling.

To favor computational stability, all variables, both dependent and independent, were linearly scaled to values between 0 and 1, before calculating the coefficients of the regression model. The minimum value of each of the variables was set to 0, while the maximum value was set to 1. All other values were scaled accordingly. For example, the coverage for high risk adults ranged from 10.4% to 47.2%. We will refer to the difference between the maximum and the minimum values in a variable as the range of the variable, in this case: $47.2 - 10.4 = 36.8$ is the range of the high-risk adult's coverage. Then, when scaling, 10.4 becomes 0, and 47.2 becomes 1, and all other points in this variable take values according to their difference with the minimum value in proportion to the range.

In this case, a coverage of 30% would become after scaling: $(30 - 10.4)/(47.2 - 10.4) = 0.53$. Following the same example, the values for the coverage in children ranged from 21.3% to 84.7% before the logarithmic transformation. Then, the transformed values range from $\ln(21.3)$ to $\ln(84.7)$. This is, from 3.059 to 4.439. Scaling these values between 0 and 1, the minimum 3.059 becomes 0, while the maximum 4.439 becomes 1. In this case, a coverage of 30% for children would be transformed to $\ln(30) = 3.401$, and this value would become $(3.401 - 3.059)/(4.439 - 3.059) = 0.248$ in a 0-1 scale.

Since all of the variables were scaled between 0 and 1 before calculating the linear regression coefficients, each of the coefficients represents a proportion between the expected change (in terms of the range) of the dependent variable and a change (in terms of the range) of its independent variable, given that all other variables remain unchanged. Since the values of the high-risk model were only scaled (and not transformed), this interpretation extends directly to the un-scaled (real) values. As an example, the coefficient for the previous seasonal influenza coverage has a value of 0.36, and the seasonal influenza variable has a range of 43.9 (55.4 maximum minus 11.5 minimum). According to our model, if a state would increase its seasonal influenza coverage by 10% of the range for this variable (a percent change of 4.39% in the proportion of people vaccinated for seasonal influenza) it would expect to increase by $0.36 * 0.1 * (\text{range of dependent variable})$, which equals $0.36 * 0.1 * 43.9 = 1.58\%$ the high-risk adult 2009 H1N1 vaccination coverage, if all other variables remain unchanged. Notice that we can address the effect of a change without necessarily considering the initial value for the variables. This is possible due to the linearity of the model.

For the children's model, the logarithmic transformation requires a more careful interpretation. The dependent variable of the regression model is the logarithmic transformation of the real dependent variable, the 2009 H1N1 vaccination coverage for children. Given that the transformation is logarithmic and not linear, we need to consider the value of the dependent and the independent variables, as well as the change, to estimate the new value of the true dependent variable. For example, the coefficient for the % of children population is -0.18817. The range of this variable is 22.1, and the range of the logarithmic transformation of coverage is 1.38. Suppose that a certain state had an H1N1 coverage of 30% for children, with a 10% of its population younger than 18 years. In this example, an assumed increment of 2.21% in the proportion of children population (10% of its range) would cause a decrease of 1.88% of the range of H1N1 coverage for children in that state, if all other variables remain unchanged. The transformation of the 30% coverage is $\ln(30) = 3.401$, and the size of the change in this transformed variable is $-0.0188 * 1.38 = -0.026$. Therefore, the final value of the transformed variable is $3.401 - 0.026 = 3.375$. The real value of the coverage will be $\exp(3.375) = 29.2\%$, representing a real change in H1N1 coverage of only 0.8%.

One of the main assumptions to achieve optimality of the coefficients' estimation in linear regression is that the distribution of the dependent variable, conditional on the independent variables, is normal. In other words, the error term of the regression model is assumed normal. This assumption does not imply normality of the independent variables, which is not required for linear regression. Additionally, since normality of the conditional distribution of the dependent variable is rarely found, approximation to

normality is usually accepted, understanding that such approximation will provide good coefficient estimators, but likely not optimal (59).

Results

Nine independent variables were significantly associated with vaccination coverage in children and eight for high-risk adults (fifteen different independent variables in total, two of which are shared by both models). A list of these variables can be found in Table 3.1. The adjusted R-squared for the regression models is 0.82 for children (Table 3.2) and 0.78 for high-risk adults (Table 3.3), and both of their p-values are close to 0.

Table 3.1: List of variables appearing in either model, including the dependent variables at the top. Table shows the variable's name, description, reference for the data, average (Avg), standard.

Variable	Description of Variables	Reference	Avg	S.D.	Max	Min
D-1	Coverage of Children 6 months to 17 years	MMWR(7)	38.9	11.9	84.7	21.3
D-2	Coverage of persons aged 25--64 years at high-risk	MMWR(7)	25.4	7.6	47.2	10.4
I-1	Percent of Women Age 18 and Older Who Report Having Had a Pap Smear Within the Last Three Years, 2008	State Health Facts(25)	82.7	2.9	88.9	74.1
I-2	Maximum number of vaccination sites per state per thousand population (2009)	CDC Report (Calculation) (3)	0.5	0.1	0.7	0.00*
I-3	Percentage Reporting Not Seeing a Doctor in the Past 12 Months Because of Cost	State Health Facts(25)	13	3.4	20.5	6.2
I-4	Underserved Population Living in Primary Care Health Professional Shortage Areas, as of September, 2008	State Health Facts(25)	12.6	7.6	34.4	1.7
I-5	Resident population under 18 years, percent (July 1 - estimate) 2008	Census(17)	24	1.9	31	18.9

Table 3.1 (continued)

Variable	Description of Variables	Reference	Avg	S.D.	Max	Min
I-6	Resident population: American Indian and Alaska Native alone, percent (July 1 - estimate) 2008	Census(17)	1.8	2.9	15.3	0.2
I-7	Total Public Doses Oct-Feb divided by Estimated People Vaccinated	CDC Report(35)	39.6	20.3	98.9	11.9
I-8	H1N1 Vaccine Doses Distributed or Administered to Date from Large Pharmacy Chains / Retail-Based Clinics to States as of January 29, 2010	CDC Report(34)	10	6.6	30.1	0
I-9	Seasonal influenza Coverage for adults 18 -- 49 yrs on the 2007-2008 season	CDC Influenza Vaccination Coverage(55)	55.4	11.5	80.5	27.3
I-10	Natural logarithm of the ratio of number of shipments to number of ship-to-sites	CDC Report(2) (Calculation)	4	3	19.6	1.6
I-11	Percentage of doses sent to primary care, MDs, counties, hospitals, urgent care, clinics, or pharmacies.	CDC Report(2) (Calculation)	66.4	20.1	99.4	4.7
I-12	Total number of cars, trucks and buses per capita	National Auto Dealers Association(21)	0.81	0.16	1.19	0.36
I-13	Coverage expanded to general population by Dec. 4, 2009: 1- Yes, 0- No	University of Michigan Summary(32)	0.42	0.49	1	0
I-14	School-based clinic strategy: 1- School Focus, 2- No	University of Michigan Summary(32)	0.3	0.46	1	0
I-15	Third parties make transfer/redistribution to small providers	University of Michigan Summary(32)	0.16	0.37	1	0

For children, four factors related with supply chain and campaign processes contributed positively to coverage: average ratio of the number of shipments per ship-to-sites, the state focus on school vaccination, the use of third parties (i.e. state or locally hired distributors) for further distribution to small providers, and the estimated proportion of doses that were administered in public sites.

Table 3.2. Regression results for predicting the state level vaccination coverage for children 6 months – 17 years, United States, end of January 2013.

Coefficients when predicting ln(children coverage percentage)					
Variable	Short Description	Estimate	Std. Error	t-value	Pr(> t)
	(Intercept)	0.01488	0.05956	0.25	0.804
Indep15	(Re)Shipments	0.42308	0.07285	5.807	0.000
Indep19	Focus on School	0.36769	0.07239	5.079	0.000
Indep6	Max # Sites	0.29734	0.07016	4.238	0.000
Indep20	3rd Party Dist'n	0.24461	0.06349	3.852	0.000
Indep12	% Public Doses	0.2125	0.06837	3.108	0.003
Indep10	% Children	-0.18817	0.07965	-2.362	0.023
Indep17	Cars per capita	-0.2843	0.07726	-3.68	0.001
Indep9	% Underserved population	-0.28992	0.07701	-3.765	0.001
Indep8	% Visit, Cost	-0.35139	0.08217	-4.276	0.000

Two factors were related to existing health infrastructure: the maximum number of ship-to-sites had a positive association with coverage, and the percentage of medically underserved population (proportion of population living in primary care health professional shortage areas (25)) a negative association. Coverage was also negatively associated with population factors including the percentage of the population that will not visit a medical doctor because of cost, the number of vehicles per capita, and the percentage of population under 18 years old.

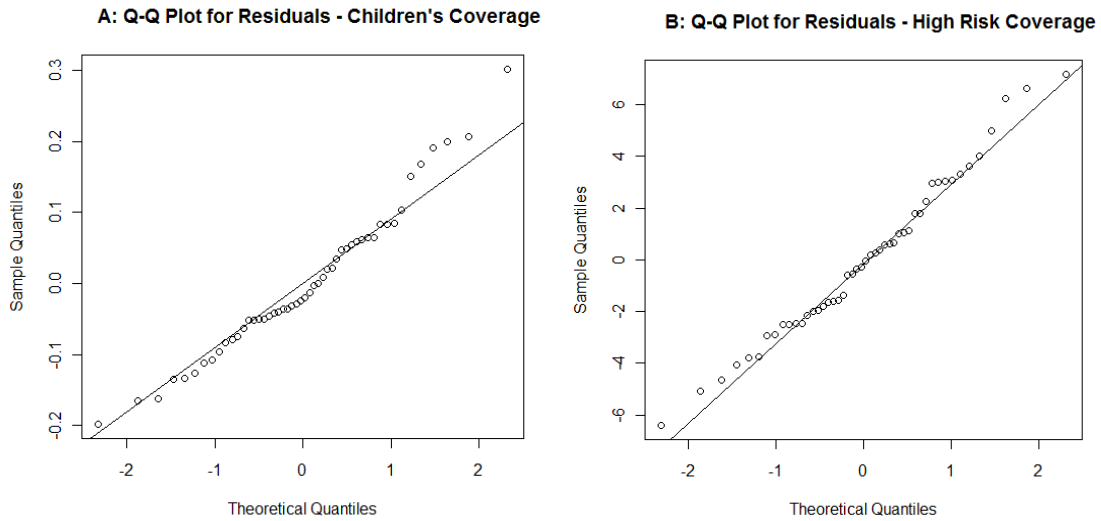
Table 3.3: Regression results for predicting the state level vaccination coverage for the high-risk adult population, United States, End of January 2010.

Coefficients when predicting coverage of high-risk adults					
Variable	Short Description	Estimate	Std. Error	t value	Pr(> t)
	(Intercept)	-0.46318	0.09916	-4.671	3.51E-05
Indep1	Women with Pap Smear	1.44641	0.38477	3.759	0.000559
Indep6	Max # Sites	0.54139	0.09705	5.579	1.99E-06
Indep16	% Doses shipped to “general access locations”	0.38443	0.07088	5.424	3.26E-06
Indep14	Previous seasonal influenza coverage adults	0.3603	0.08525	4.226	0.000138
Indep11	Pop. American Indian	0.20897	0.0777	2.69	0.010474
Indep13	Pharmacy and Retail	0.17915	0.05251	3.412	0.001515
Indep18	Expanded by Dec. 4th	-0.11829	0.02398	-4.933	1.55E-05
Indep9	% Underserved population	-0.37442	0.06081	-6.157	3.14E-07

For high-risk adults, two supply chain processes were positively associated with uptake: the percentage of doses shipped to “general public” locations, and the use of pharmacy and retail locations for vaccination; and one, the expansion of vaccination to the general public by December 4th, was negatively associated. Coverage was positively associated with population and health related factors: percentage of women with a Pap smear, past seasonal influenza vaccination, and percentage of population that is American Indian. Two infrastructure factors were associated: the proportion of the population

medically underserved (negatively) and the maximum number of ship-to-sites (positively).

Figure 3.3A presents the Q-Q plot for the residuals of the children’s model, and Figure 3.3B the Q-Q plot for the residuals of the high-risk adults’ model. The closer the plotted points are from the straight line, the better its distribution approximates normality.



Figures 3.3A and 3.3B show Q-Q plots for the residuals of the children’s and high-risk adult’s models, respectively.

Discussion

We sought to identify factors related to vaccination program decisions and processes that may have facilitated or hindered vaccine uptake for two target groups for vaccination: children and high-risk adults. Several supply chain and system factors were associated with vaccination coverage of children and of high risk adults. With the

exception of the maximum number of ship-to sites, a factor that was also associated with overall adult coverage [3], factors associated with coverage of children and of high risk adults did not overlap. Additionally, factors not related to program decisions such as health-seeking behaviors and population characteristics were also associated with state-to-state variation, as would be expected given baseline variation in vaccination coverage for recommended vaccines(25) and the variety of factors associated with vaccinations, both for high-risk individuals(48, 50, 51, 60) and children(40, 42).

Several findings were related to the type of providers or locations to which vaccine was directed. For children, having a focus on school vaccination was associated with higher coverage (five of the six states that achieved the highest coverage in children implemented statewide school vaccination programs(7, 32)), as was distribution to public sites. Public sites can include schools, but also locations such as mass clinics run by health departments. For high-risk adults, more distribution to providers with a broad base of access (including pharmacies, primary care providers, county health departments, etc.) was associated with higher coverage. It is noteworthy that coverage for overall adults was negatively associated with distribution to internists or specialists [3]. Taken together, the results for adults suggest that vaccine that was broadly accessible may have facilitated higher coverage. This could be because high-risk adults may not visit internists or specialists frequently enough to be vaccinated in this time period; that specialists traditionally have had less focus on vaccinating so patients may have looked elsewhere for vaccine, or that the cost in some settings was lower. For high-risk adults, the percent medically underserved is also negatively associated with coverage, which may also help explain the positive impact of open access locations and pharmacies.

The number of shipments per ship-to site was positively associated with coverage for children but not for high-risk adults. For children, this may reflect repeated shipments to locations such as local health departments, mass clinics, or pediatricians who may have offered repeated clinics. Some health departments monitored usage and distributed more vaccine to providers who were depleting vaccine supply faster, which is another potential hypothesis. The maximum number of sites to which vaccine could be directly shipped through the centralized distribution system was positively associated with vaccination coverage for both children and high-risk adults, a finding also observed for overall adults [3]. Because the number of ship-to-sites allowed for each state was based on a formula that included the population size as well as the number of existing VFC providers, this measure may reflect a more robust healthcare infrastructure.

The expansion of vaccine availability to the general public by December 4th was associated with lower coverage for high risk adults. Early expansion could have resulted in less access for high-risk adults, especially if a state had sequential priorities (e.g., children first, then high-risk adults). However, because in most states, decisions about when to make vaccine available beyond the initial target groups were based on perceived demand for vaccine, e.g., as ascertained from provider vaccine orders and attendance at public clinics, so the decision to expand early could reflect lower demand in those states.

Coverage for high risk adults was positively associated with uptake of seasonal vaccine for high-risk adults in 2007-2008, as it was for adults overall (47). This could be because the administration sites for adults were similar to past seasonal influenza campaigns or it could reflect use of preventative services. In contrast, the lack of association for children could reflect the fact that vaccine administration sites differed

from past seasons with school vaccination playing an unprecedented role during this influenza vaccination campaign. A second hypothesis for children is that the increased focus on them as a priority group served to motivate their vaccination by caregivers or providers. The association between coverage rates in high-risk adults and rates of receipt of Pap smear may be a reflection of utilization of preventive care in a state, and could also reflect vaccination by Ob-Gyns (61).

For children, lower coverage was associated with a higher percent of the population reporting they would not visit a medical provider because of cost; and coverage was positively associated with the proportion of vaccine being directed to public sites. These findings may relate to the relationship between cost and access (e.g., a mass clinic may have been free to patients, while visiting a specialty physician may result in a fee), as we found for high-risk adults. It is noteworthy that for both children and high-risk adults, the percent uninsured was highly correlated with coverage (though it did not add to the model).

The negative association between coverage for children and the percentage of the population under 18 could be a combination of the pro-rata allocation and prioritization policies. Given the initial focus on vaccinating children, the amount of vaccine available per child was less in states with proportionately more children. Additionally, the vaccine available per child decreased since a second dose was recommended for children 6 months through 9 years of age(62). In the event of a vaccine shortage, deviating from an overall pro-rata allocation may be justifiable, if a sub-population at higher risk is easy to identify, and the impact of increased allocation to this sub-population is potentially large.

This warrants further examination given the complexity of recommendations with multiple target groups.

The use of third party distribution and number of cars per capita appeared in the model for children. Both have small individual correlations with the dependent variable, so they improve the overall model fit when controlling for other variables.

This study had several limitations. As explained more fully in the article by Davila-Payan et al.(63) the shipment data ends December 9 2009, but we examine vaccination coverage at the end of January 2010. We also do not know where the vaccine was actually administered; this means for example, that we do not know whether repeated shipments to the same location, i.e., a local health department, were being distributed through mass clinics, schools, or other local providers. We were only able to determine provider type for 75% of shipments, and the information on state and local decisions and processes was not always complete. Modeling limitations include the fact that ecological approaches do not point to individual characteristics of the population but to state-level conditions, leaving out potentially relevant variations within states, and that that cross-sectional studies cannot determine causality. Also related to the latter, it should be noted that there are multiple potential explanations for findings. While we aimed to include the most likely ones, the potential for bias should be recognized. Additionally, we are identifying associations with a relatively small number of dependent variables (51), across many independent variables that have correlations, and confidence intervals of the coverage estimations were not considered in the regression. We have kept the best models we found, however, other good models could also exist. The supplementary table in Appendix C presents a summary of variables highly correlated with those in the

children and high-risk models. Our models provide a solid approach on the analysis of factors related with coverage. However, care should be taken in relying too heavily on any particular variable or finding without considering its interaction with other variables in the model.

The distribution and administration of the H1N1 vaccine provided an opportunity to understand how specific approaches may affect vaccine uptake in priority populations in an emergency situation. Results from this analysis complement those examining factors associated with vaccination of overall adults and suggests that supply chain factors may affect vaccine uptake. The analysis also points to opportunities for future research such as further analysis on uptake and the relationship with spatial access to vaccine or access by provider type, and the role of urban or rural differences in vaccine uptake. These research questions and others can be informed by more detailed mapping of the process and system to show details of demand (e.g., by population or providers), supply (e.g. details on allocations and shipments including the final point of distribution and the category of provider), lead-times across the system, variations within and across states, where vaccine was administered, when, by who and to what subpopulation. Such data would also allow for a robust comparison of potential distribution systems and processes before they are implemented.

CHAPTER 4

ESTIMATING CHILDHOOD HIGH-BMI BASELINE PREVALENCE IN SMALL GEOGRAPHIC AREAS

Introduction

Obesity is one of the most urgent health challenges considered by the Centers for Disease Control and Prevention (CDC) to impact health indicators in the US, and was included on their recently released list of “winnable battles”(64). Studies show that childhood obesity in the US remains at its highest point in history (65), with an approximately 16.9% prevalence in the 2–19 years old US population in 2010.

Overweight (at or above 85% and below 95%) and obesity (at or above 95% of the BMI-for-age growth charts) (66, 67) in children pose both present and future health risks (68), making this population a target for interventions aiming to improve health of the overall population or children specifically. Interventions can include: inducing healthy behaviors (69), changing activity and inactivity patterns (70), modifying eating habits (71), and promoting prevention through education programs for new parents (72). However, if interventions are targeted as a function of prevalence, then limited resources can be allocated proportional to the need. Targeting interventions geographically is one strategy, and thus identifying areas with children at higher risk for elevated body mass index (BMI) is crucial in delivering cost-effective interventions. Our study considers high BMI children those who are either overweight or obese.

Existing weight and height data used for the estimation of children's BMI levels in the US has been obtained either through self-reports (which include parental reports), or through direct measurements on the surveyed individuals. For example, the National Survey of Children's Health (NSCH) (73) provides state-level percentages of overweight or obese youths age 10-17 from self-reported information. The National Health and Nutrition Examination Survey (NHANES)(74) presents weight and height data from direct measurements on the sampled population, from which national estimations of childhood overweight and obesity can be calculated (75). The Youth Risk Behavior Surveillance System (YRBSS) estimates national, state, and large urban school district obese and overweight prevalence values based on self-reported height and weight in representative samples of students in grades 9–12 (76).

For a number of reasons (e.g. displacement of human resources, use of measuring devices, etc.) estimation of high BMI prevalence through direct measurement is more challenging and costly than through questionnaire surveys, which can be performed remotely. On the other hand, self-reporting may induce additional inaccuracies in estimation of high BMI prevalence. Self-reported information is generally biased in groups of children younger than 12 years old (77), and there are no reasonable approaches for correcting this bias for accurate estimation of high BMI prevalence (78). For these young groups, direct measurement may provide the only reliable method to estimate high BMI prevalence (78). Therefore, estimating the BMI status for children in geographic areas without available direct measurement is a challenging problem.

Some cities or states have begun initiatives to measure BMI in schools (79). In general, systemic measures from schools or local random samples are not available for

most small geographical areas, such as counties, school districts, zip codes, or census tracts, to use in the estimation of their high BMI prevalence. Researchers can obtain geocoded data from NHANES after undergoing a review process (80); however, there is no direct access to these data for most stakeholders in public health and not all small geographical areas are included in their sampling.

Because of the difficulties in using self-reported BMI data, and the shortage of small-area samples for children, the existing methodologies used to estimate (adult) obesity prevalence in small areas (81-84) cannot be applied. To address this limitation, we provide an approach that uses publicly available information for baseline estimation of high BMI prevalence in children; we accompany our point estimates with confidence intervals. Our prediction model is similar to some found in (85); specifically, we build a logistic regression model using individual survey data and complement it with a re-sampling procedure for estimating the distribution of the model coefficients. The logistic regression is estimated using information for high BMI covariates in children 2-17 years old from the 2001–2010 continuous NHANES (74). This model is used to simulate the BMI status of individuals in a population by virtually generating population samples using data from the 2010 Census Summary File 1 (86) and specific searches in the 2010 American Community Survey (87).

Using publicly available data in the construction of this methodological approach may offer a wider opportunity for its application by others, including not only public health organizations but also stakeholders like healthcare providers. Since our approach is presented in the absence of local samples or information on existing interventions, we

call our estimations *baseline prevalence estimates*, in the expectation that they will eventually be replaced by estimators that include local sampling through direct measures.

Approaches to the estimation of prevalence of certain health conditions, in small geographic-areas, and without the benefit of local sampling are not new. Usually, these approaches are split in at least two phases: building a generalized linear regression model from a representative survey, and then connecting the regression model to a specific population through its socio-demographic characteristics. Some of the main difficulties researchers encounter with the use of this type of approach are: Correctly capturing the distribution of the regression coefficients into the model, restrictions in the aggregation of the publicly available socio-demographic information in the small geographic areas, generating adequate confidence intervals for the prevalence estimation, and validation of the approach due to lack of local samples to compare to. For example, Messer et al. (88) benefited from being granted access to records with the detailed characteristics of the target population, but do not capture the distribution of the regression coefficients, and their confidence intervals are roughly estimated. Some use the geographic identifiers of surveys to obtain local samples (89), although these identifiers are usually of restricted access. Using geographic identifiers has become a popular approach for the estimation for adult obesity in small geographic areas (81, 82, 90). Some have used Bayesian simulation based software to estimate the uncertainty of their estimations (91, 92).

Choy et al. (93) stress the relevance of using publicly available information, and build an estimation model for disease prevalence combining NHANES III and Census 1990 data. Their results are validated showing significant correlations to related malignances, although the paper does not provide insights on the use of more recent (and

aggregated) census data or on the estimation of variability for the estimation. Other studies validate the use of generalized linear models to extrapolate national surveys into small geographical areas using Census 2000 data (94), but do not provide any guidance on how to deal with the common difficulties implied.

Zhang, et.al(95) present a multilevel approach that uses restricted geo-coded data from the 2007 National Survey of Children's health to predict childhood (ages 10 to 17 years) obesity at the census block-group level. But, as noted by Longjohn et.al. (96), a transcending effort needs to be made to provide local BMI data on children 14 and below.

We predict high BMI prevalence for children within a small geographic region by simulating its virtual population at the individual level and predicting whether each individual has high BMI. We use the conditional distribution of high BMI probability derived from the regression model for the construction of the confidence intervals, and extract all data from publicly available sources. Therefore, our approach is a novel methodology for the estimation of baseline prevalence in small geographic areas without the use of restricted geographic identifiers and in the absence of available local samples.

The development of our approach responded to the need of prevalence estimation for targeted interventions of a health care provider. This institution desired information to target a large scale campaign to improve children's health conditions in Georgia, including interventions addressing the state's childhood high BMI problem. This paper explains how the methodology is used to provide baseline estimations for the high BMI prevalence of their target population (children 2-17 years old) in Georgia, at the census tract level. Nonetheless, the same methodology can be applied to also generate baseline estimators at zip code, county or other geographic aggregation levels. With prevalence

mapped at the local level, we are able to identify areas where the greatest impact could be made in the state (based on prevalence and overall number of high BMI children), and a deployment strategy for the intervention is proposed.

Below we present a more detail explanation of our methods, followed by the presentations of the results derived for the state of Georgia. Finally, we will discuss the application, limitations, and advantages of our approach.

Methods

We develop a local-level high BMI prevalence model using NHANES continuous surveys (74). In this process, we merged data from 2001 to 2010. Then, we used 2010 Census data to generate virtual populations. The model implementation uses R statistical software (58) to read the data files(97), manipulate data, choose the set of initial census-based predictors, and generate samples from the distribution of a high BMI event based on the fitted logistic regression model(98), and to map the results(99). It also uses the C++ platform to simulate the prediction of high BMI prevalence at the census tract level. Below we present details on this methodology.

Logistic Regression Model

In fitting the logistic model, we observe $(X_i; Y_i)$ for $i = 1, \dots, N$, for N individuals, where Y_i are binary values specifying whether the i^{th} individual has high BMI and $X_i = (X_{i_a}, \dots, X_{i_h})$ is the set of covariates for the i^{th} individual. The regression problem is to estimate the probability of an individual to have high BMI given their characteristics represented by the set of predictors X_i , specifically, estimate $\Pr(Y_i=1 | X_i)$. The descriptions of the binary response Y and the set of covariates X are provided next.

As a comparison, we conducted an additional analysis using linear regression, which we present in Appendix D.

Derivation of the dependent variable: Y

We use the conventional calculation of the BMI as a person's weight in kilograms divided by the square of their height in meters. The BMI-for-age charts adopted by the Centers for Disease Control and Prevention (CDC) in 2000 define our population percentiles (66). The present work considers that individuals have high BMI when their BMI reaches or exceeds the 85th percentile. NHANES measures individual BMI values in the examination of interviewees. We therefore convert the BMI values into a vector consisting of binary values (equal to 1 when the individual has high BMI and 0 otherwise). This variable is the dependent variable in the regression model.

Model Covariates: X

Based on the findings in the existing literature (100, 101), we use covariates related to socio-economic status and a number of other individual factors potentially related to high BMI. Some of the individual-characteristics variables reported by NHANES between 2001 and 2010 are also found in the publicly available information for census tracts in the 2010 Census. We consider variables common to both datasets and expectedly associated with high BMI for building potential covariates in the model. Examples of the variables considered are: gender, race/ethnicity, age, education level of the household reference person (person who owns or rents the residence where the members of the household reside(74)), household size, and income or poverty level. Table 1 summarizes the covariates considered. We test all the covariates in the iterative building of prospective models. We remove from the models those variables that are not

statistically relevant or not closely related to the publicly available census information.

Table 4.1 summarizes the covariates considered.

Table 4.1: Summary of covariates tested in logistic regression model

Covariate	Type	Values	Meaning
X_b	Binary	{0, 1}	Non-Hispanic Black (1) or not (0)
X_{nho}	Binary	{0, 1}	Non-Hispanic Other (1) or not (0)
X_h	Binary	{0, 1}	Hispanic (1) or not (0)
X_4	Binary	{0, 1}	Below 4 times poverty level (1) or not (0)
X_2	Binary	{0, 1}	Below 2 times poverty level (1) or not (0)
X_1	Binary	{0, 1}	Below poverty level (1) or not (0)
X_g	Binary	{0, 1}	Male (1) or Female (0)
X_e	Discrete	{1, ..., 5}	Household representative education level (lesser to higher)
X_h	Discrete	{2, ..., 7}	Household size 2 to 7 or more
X_i	Discrete	{1, ..., 11}	Household income level (Increasing)
X_a	Continuous	(24, 216)	Age in months

Implementation of the logistic regression model

We linearly scale all variables into a [0,1] interval for numerical stability and comparison across covariates. We select the covariates using backward stepwise variable elimination. The final model includes covariates that: 1) significantly explain the variability in the response variable (p-value < 0.02); and 2) can be obtained through publicly available census tables (see complete list of census tables used in Appendix E).

The fitted logistic regression provides estimates for the conditional distribution of $Y_i | X_i$. The estimated model for $\Pr(Y_i = 1 | X_i = x)$ is:

$$\hat{p}_i = H\{b_0 + b_a x_{i a} + \dots + b_h x_{i h}\}$$

where H is the logistic regression function and b_0, b_a, \dots, b_h are estimated regression coefficients. In order to obtain a sample from the empirical distribution of \hat{p}_i , or equivalently, from the empirical distributions of b_0, b_a, \dots, b_h , we use the bootstrap re-sampling method. Specifically, we repeat for 1000 times the following procedure:

Divide NHANES data in ten random classes with equivalent number of samples in each.

1. Removing one class at a time, fit the logistic model to the remaining sample of individuals to obtain a realization $b_0^*, b_a^*, \dots, b_h^*$ from the empirical distribution of b_0, b_a, \dots, b_h . The sampled probability from the empirical distribution of \hat{p}_i becomes $\hat{p}_i^* = H\{b_0^* + b_a^* x_{i a} + \dots + b_h^* x_{i h}\}$.

We use the sample from the empirical distribution of \hat{p}_i in predicting the high BMI probability of a virtual individual which in turn is used to estimate high BMI prevalence within a geographic area based on its population composition.

Simulation Model

Generating virtual population with a geographic area

Given a geographic area such as a census tract, we can obtain demographic and socio-economic data using publicly available data from the Census Bureau. These data can be used to generate a virtual population within that specific geographic area. This is equivalent to generating from the distribution of $X_i = (X_{i a}, \dots, X_{i h})$. In our simulation study, we obtain 1000 virtual individuals within each census tract by simulating from the distribution of population characteristics X . The generated characteristics of these

individuals are denoted by X_1^*, \dots, X_{1000}^* . In our implementation, we focus on the census tract as our primary small-area of interest, although the methodology presented herein can also be used to analyze other unit areas including county or zip code.

To be able to have a closer characterization of the population in a geographic area, we consider the possible interdependence of some of the population characteristics X_{i_a}, \dots, X_{i_h} . Specifically, we relate each of the census variables to the race/ethnicity groups by examining multiple tables provided by the Census. We find that, for example, the distribution in the household size for non-Hispanic Whites is different than that of Hispanics. Other interactions are not publicly available in the Census, and therefore, conditionally on race/ethnicity, the other variables are simulated independently of each other.

Linking the individual high BMI regression model to small-area-level data

Our simulation generates a virtual sample of the population living in a geographic area using census information and accounting for some of the interdependence in the population characteristics. Specifically, for each virtual individual j ($j = 1, \dots, 1000$ in our implementation) we obtain \hat{p}_j , the estimate of $\Pr(Y_j = 1 \mid X_j = X_j^*)$ by sampling from the empirical distribution of b_0, b_a, \dots, b_h , resulting in the realization $b_{0j}^*, b_{aj}^*, \dots, b_{hj}^*$, and evaluating the logistic probability function in X_j^* , the personal characteristics of the j -th virtual individual, calculating

$$\hat{p}_j^* = H\{b_0^* + b_{a_j}^* X_{a_j}^* + \dots + b_{h_j}^* X_{h_j}^*\}.$$

To label each individual as high BMI or not, the probability of the j -th virtual individual is simulated from a Bernoulli random variable with probability \hat{p}_j^* , denoting the resulting label Y_j^* . The high-BMI prevalence estimate is then

$$\hat{p} = \frac{\sum_{j=1}^B Y_j^*}{B}$$

for $B=1000$ samples in our implementation. For each census tract, we also estimate the $1-\alpha$ confidence interval for the high-BMI prevalence by repeating the simulation approach 1000 times to obtain $\hat{P}_1, \hat{P}_2, \dots, \hat{P}_{1000}$ samples from the distribution of high-BMI prevalence within the census tract and estimate a normal interval as

$$\left(\bar{P} - z_{\alpha/2} \sqrt{\hat{V}}, \bar{P} + z_{\alpha/2} \sqrt{\hat{V}} \right) \text{ where } \bar{P} = \frac{\sum_{s=1}^{1000} \hat{P}_s}{1000} \text{ and } \hat{V} = \frac{1}{999} \sum_{s=1}^{1000} (\hat{P}_s - \bar{P})^2.$$

It is important to remember that the sampling is performed on virtual individuals that depict the characteristics of the population and not on population members.

Therefore, 1000 independent samples can be made even if actual communities have a children population of less than 1000 individuals.

This simulation model allows for variations in the predicted probability \hat{p}_j^* due to estimation by using a realization of the estimated regression coefficients and due to individual randomness by simulating individual characteristics. The confidence intervals are estimated using a double re-sampling technique for more accurate estimation of the variance of the prevalence estimate.

Application to Targeted Interventions

When a limited amount of resources is assigned to improve an overall indicator of a system, a possible solution will be to allocate most of those resources to those components of the system that will render the largest overall benefit to that indicator. For our context, two indicators are relevant for the priority classification of small areas based on the severity of the childhood high BMI problem. The first is the estimated baseline

prevalence for the area as calculated. The second is the estimated number of high BMI children (the product of the estimated prevalence times the population of children) for each area.

The commonly known Pareto principle(102) (which has to be taken with extreme care when used in relation to health) establishes the idea that when looking for the solution of the problems of a system, usually the largest proportion of the benefit is obtained from solving a proportionally small part of the problems. We use a variant of this Pareto-principle to select priority areas with the goal of selecting the census tracts with a larger number of high BMI children. We prioritize those census tracts that in total represent about 80% of the total population of high BMI children.

Results

Logistic Regression Model

The logistic regression model we present includes six variables, three variables for encoding four race/ethnicity categories (Hispanic, Non-Hispanic Black, Non-Hispanic White, and Non-Hispanic Other), age of child in months, household size, and education level of household representative. The reference group for the regression is Non-Hispanic White in a household of size 2, age 2 years, and a household representative with lowest education level. The odds ratios (ORs) of the coefficients and the 95% confidence intervals of the ORs are shown in Table 4.2. All the p-values of the variables are less or equal to 2% and the p-value of the overall model is close to zero for all tests applied, including Wald F and Wald chi-square.

As shown in the model and as found in NHANES data, Non-Hispanic Black children and Hispanic children are more likely to have high BMI than the reference

group, and Non-Hispanic Others are less likely. The model also indicates that the probability of having high BMI increases with age, and decreases with household size. High BMI is also less likely for children whose household reference person has higher education.

Table 4.2: Adjusted coefficient estimates of the model for individual prediction (values of covariates were scaled to a [0,1] interval before regression) – Children 2-17 years old in Georgia.

Coefficient	Covariate	Estimate	95% C. I.	Pr(> t)	Characteristic
β_0	<i>Intercept</i>	-0.444	(-0.683, -0.205)	<0.001	
β_b	X_b	0.228	(0.115, 0.341)	<0.001	Non-Hispanic Black or not
β_{nho}	X_{nho}	-0.239	(-0.420, -0.058)	0.011	Non-Hispanic Other or not
β_h	X_h	0.342	(0.215, 0.469)	<0.001	Hispanic or not
β_e	X_e	-0.612	(-0.788, -0.437)	<0.001	Household representative education level
β_{hs}	X_h	-0.642	(-0.843, -0.440)	<0.001	Household size
β_a	X_a	0.613	(0.478, 0.748)	<0.001	Age in months

Simulation Model

The application of the simulation model is illustrated for the state of Georgia. In Table 4.3 we present the results for some counties and a few of their census tracts. It can be observed, for example, that census tract 20300 in the DeKalb County has a prevalence (0.27) that is significantly less than that of tract 20600 (0.36) in the same county (both are marked in the table), even though these census tracts are neighboring. Census tract 20300 has a different ethnic composition from tract 20600 (84%, 4%, 4%, 8% and 8%, 89%, 1%, 3% for non-Hispanic White, non-Hispanic Black, Hispanic, and Other non-Hispanic,

respectively); comparable Family HH size (average 3.0); higher education of household representative (average 4.2 versus 3.7) and lower average age of children (7.5 versus 8.7 years). Also, Table 4.4 contains some examples of the prevalence for counties in Georgia. Differences among counties are observable, but not as strong as those for census tracts. For example, Early County has a higher mean estimation of baseline prevalence (0.37) than Oconee County (0.32).

Table 4.3: Examples of census tract prevalence estimates in Georgia, with 95% Confidence Intervals (C.I.)

County	Tract	Mean	95% C. I.
Cobb	31110	0.34	(0.301, 0.373)
Cobb	31112	0.31	(0.276, 0.346)
Cobb	31205	0.33	(0.293, 0.361)
Cobb	31206	0.32	(0.285, 0.356)
DeKalb	20300	0.27	(0.240, 0.308)
DeKalb	20400	0.29	(0.251, 0.326)
DeKalb	20500	0.38	(0.339, 0.412)
DeKalb	20600	0.36	(0.324, 0.401)
Fulton	400	0.30	(0.262, 0.332)
Fulton	500	0.31	(0.273, 0.338)
Fulton	600	0.33	(0.297, 0.369)
Fulton	700	0.36	(0.323, 0.401)
Muscogee	900	0.35	(0.314, 0.394)
Muscogee	1000	0.34	(0.300, 0.373)
Muscogee	1100	0.29	(0.255, 0.324)
Muscogee	1200	0.33	(0.295, 0.361)

Table 4.4: Examples of prevalence estimates for counties in Georgia with 95% confidence intervals (C.I.)

County	County Number	Estimate	95% C.I.
Appling County	13001	0.36	(0.320, 0.394)
Bacon County	13005	0.36	(0.320, 0.392)
Calhoun County	13037	0.37	(0.336, 0.410)
Dade County	13083	0.34	(0.300, 0.372)
Early County	13099	0.37	(0.328, 0.407)
Fannin County	13111	0.34	(0.305, 0.383)
Gilmer County	13123	0.35	(0.318, 0.390)
Habersham County	13137	0.35	(0.312, 0.385)
Irwin County	13155	0.36	(0.323, 0.393)
Jackson County	13157	0.34	(0.300, 0.373)
Lamar County	13171	0.36	(0.321, 0.390)
McDuffie County	13189	0.37	(0.329, 0.403)
Newton County	13217	0.35	(0.314, 0.385)
Oconee County	13219	0.32	(0.282, 0.352)
Paulding County	13223	0.33	(0.294, 0.369)

The estimated prevalence for high BMI at a census tract level in Georgia varies from 26% to 55%. The prevalence at a county level varies from 31% to 40%. According to the Census Bureau, census tracts are generally defined according to observable characteristics and features of the areas (103), while counties are usually larger and may include areas with more diverse characteristics, which may explain the difference in the prevalence ranges.

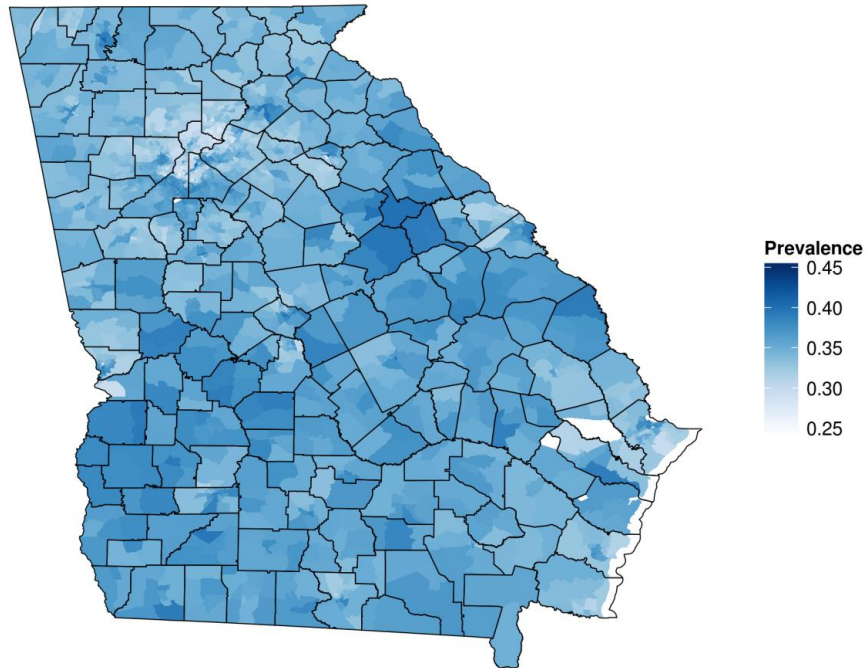


Figure 4.1: Map of baseline overweight children prevalence estimation in Georgia by census tract. Uses *R* software (58, 99)

Figure 4.1 shows the map of the baseline prevalence estimation for census tracts in Georgia. On the map, darker colors indicate higher prevalence, while lighter colors indicate a lower prevalence. It can be observed, for example, that the north area of Atlanta has estimates of low prevalence, while some areas on the east, west, and south of the city have higher prevalence estimation. This knowledge is needed in the prioritization of populations to target interventions. Figure 4.2 shows a map with the estimated number of high BMI children by census tract (more is a darker color).

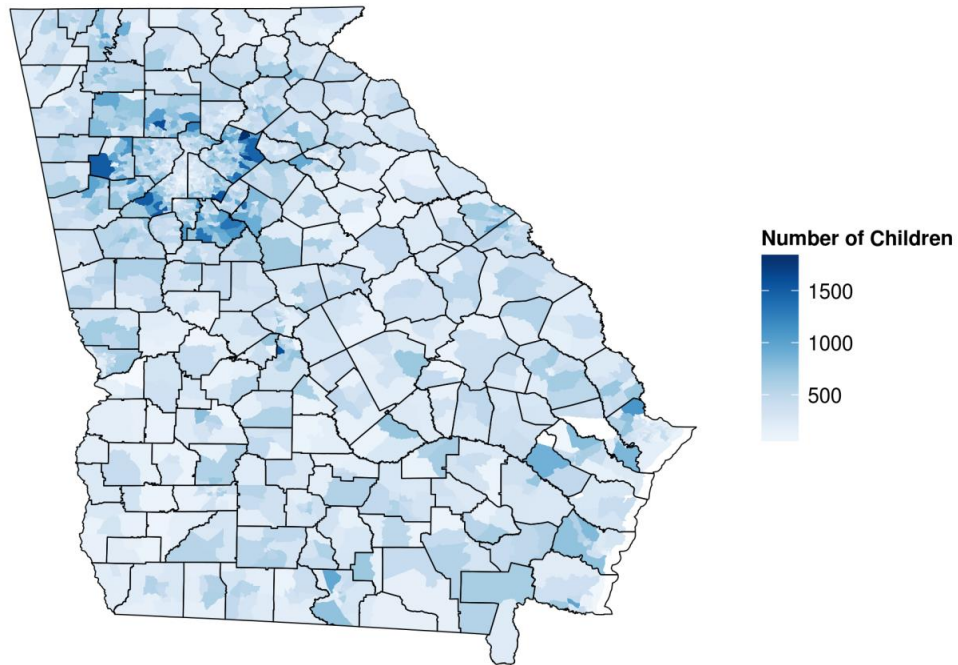


Figure 4.2: Map of overweight children population by census tract in Georgia. Uses R software (58, 99)

Application to Targeted Interventions

When classifying communities for priority in the application of interventions we identified 77% of the high BMI children live in about 25% of the counties in Georgia. Figure 4.3 shows a map of prevalence in the 39 counties, from a total of 159, which account for the majority of children with high BMI in the state. Counties are marked in darker shades when they have higher estimated baseline prevalence. This map could provide a guideline to prioritize interventions, assuming that the desired objective is to minimize the number of children with high BMI in Georgia with limited resources.

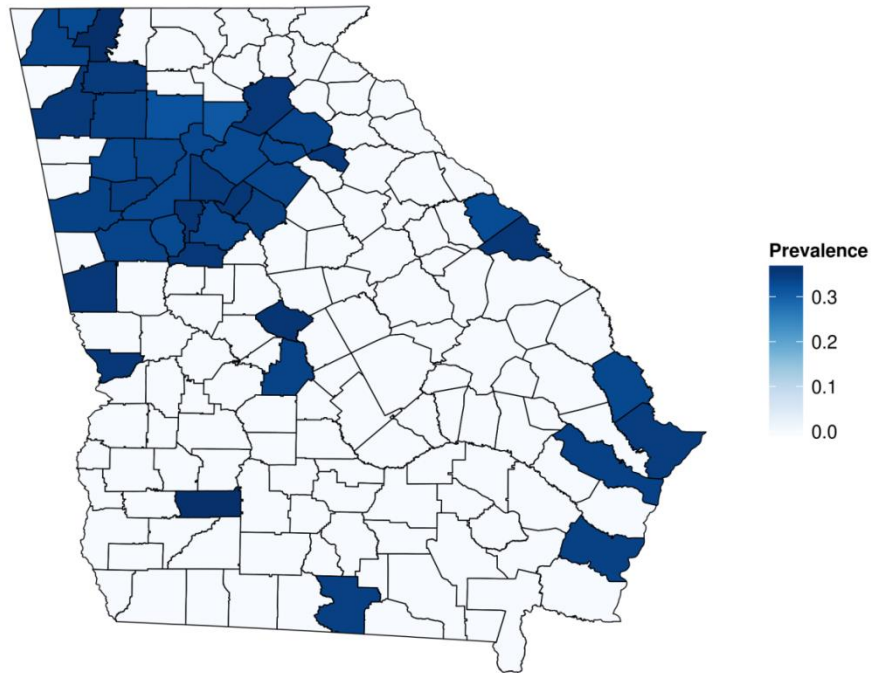


Figure 4.3: Map of overweight children population prioritization in Georgia by county. Uses R software (58, 99)

Model Validation

We develop three additional analyses to validate our modeling approach: 1) we modeled the population 10 to 17 years old in Georgia, and compared our state-level outcome to the state-level prevalence estimate of the 2007 National Survey of Children’s Health (73). 2) We modeled the population of adults and compared our county-level results with the 2007 county-level obesity estimations for Georgia by the Diabetes Data and Trends of the CDC (104). 3) We modeled obesity for children 5 – 17 years old in the

counties in Arkansas and compared to the 2006-2007 school measurements in that state (105).

Our result for the baseline prevalence estimation for high BMI children 10-17 years old in Georgia was 37.5%. This result is comparable to the 2007 estimation presented by the National Survey of Children's Health(73), which estimates the high BMI children (overweight or obese) prevalence to be 37.3% (31.7% – 42.9%) for Georgia. When comparing our county level results for adults in Georgia we obtain a 0.92 spatial correlation with the 2007 county level obesity estimations from the National Diabetes Surveillance System (104). Appendix F contains the model used for adult obesity estimation. Finally, when comparing our county level obesity outputs for children 5-17 to the 2006-2007 counties measurements in Arkansas (105) we obtain a spatial correlation of 0.78 with their estimates.

Discussion

The work presented has some intrinsic limitations. Our results do not substitute for the more complete measurement approaches available in some small geographic areas, which would also have the benefit of capturing effects of interventions. Our method is not helpful to capture the effects of locally applied interventions. Additionally, some small groups in census tracts do not have complete information available, adding levels of approximation to our approach in the smaller areas. It is difficult, if at all possible, to capture the complex nature of high BMI through modeling, which in the absence of local samples ties our approach to an unavoidable model bias. Our model is limited to the validity of the data used to build the model, and will not show the effects of changes in the population, unless they are included in the data.

Our model can only capture in a precise manner those interactions among population variables that are publicly available in the Census, introducing an increased variability into the estimation by considering independence among some of the input variables, as stated above. Many of the factors that relate to overweight in children may be area-specific and cannot be included in our model without the use of usually restricted geo-coded information.

In the paper we describe a modeling approach that combines logistic regression modeling of individuals with bootstrap resampling and simulation of virtual sampling of the population. Our approach allows us to estimate mean prevalence of overweight children as well as confidence intervals in many small areas. We built the models and populated the simulation intentionally with publicly available data so that the results could be repeated by others in the public health community. Sensible estimation differences found in a single county support the mentioned importance of estimating for small-areas. Validations indicate that our modeling methodology can provide reasonable estimates. Using our results and the proposed prioritization strategy, a local health provider in Georgia prioritized some interventions geographically across the state.

The selection of the variables in the model such as the races/ethnicities discussed in the paper could change for different regions. For our analysis of Georgia we found that three main groups comprise the great majority of the population: Black, Hispanic and White. If this study is to be conducted in any other state, the selection of the model variables should match the social composition of the region under analysis, for example by representing Native Americans or Asians.

This method represents a viable and economic alternative for estimating the baseline prevalence of high BMI children. It can also be used for the estimation of baseline obesity (BMI \geq 95%) prevalence for children or youths. Additionally, the approach can be generalized to estimate the prevalence of other diseases or conditions, especially when local samples are not available.

We present a cost effective and sound alternative to the absence of local information that can be used to inform interventions against high BMI in children based on publicly available data. With appropriate caution, the prevalence rates generated through our model have served to build maps of baseline estimation of overweight prevalence in Georgia, and can certainly help build baseline estimations for other states or diseases without publicly available small-area estimates. Finally, our methodology could be easily expanded to consider additional area-specific values if geocoded information were available. This additional information would be expected to add strength to our estimations.

Conclusion

Valid and statistically sound baseline prevalence of high BMI (or overweight, or obese) can be estimated in small geographic areas, through publicly available data, for the guidance of local interventions in the absence of direct estimations. This approach was followed for the implementation of an intervention in Georgia, and can be used by any children's health stakeholder in in the US. Future efforts include strengthening the approach by the inclusion of related covariates whose value is estimated at local levels, such as adult and youth overweight and obesity levels. Generalized regression models

allow many nonlinear forms which may capture more closely the relationship between covariates and prevalence, and can be used to strengthen this approach.

CHAPTER 5

MODELING DISRUPTIONS TO FOOD SUPPLY CHAINS IN THE U.S. CAUSED BY ABSENTEEISM DURING A PANDEMIC

Introduction

During the onset of the Influenza Pandemic of 2009 many questions were posed respect to what would be the possible effects of such event in many dimensions. One of these questions was what could be the effect of the generalized disease caused by the novel H1N1 flu virus on the supply of critical goods, such as medical materials and food. To approach this question, this chapter presents a particular analysis on the supply chain of one food type. Given its generalized consumption, its relevant place in the nutrition of people of all ages in the U. S., particularly children, and its relatively traceable supply chain, drinkable milk was selected as an instance of food supply chains for the analysis.

This study analyzes the level of disruption caused by absenteeism due to pandemic influenza, or other generalized flu-like illness, on a critical supply-chain in the U. S., and its consequences to the different echelon levels in it. The analysis also presents the effects of the disruptions under different inventory and slack capacity limits, and attempts to infer on the possible consequences of a number of proposed service policy scenarios.

To perform this research we first found all possible relevant information related with milk production, processing, and consumption, using open sources on the internet. Then, with that information on hand, we built a generic supply chain. We fixed on it all

the clearly-identified detailed information and connections, and generated all the missing connections that would allow the product to flow from the producers to the markets. We use optimization to choose the arcs that would render a minimum cost (minimum distance) for that missing information. Next, we convert the completed supply chain into a network with capacitated arcs. Finally, we built a simulation of the flow of milk in the network, and used it to virtually disrupt the Supply Chain by reducing the flows proportionally to a simulated absenteeism in work groups, and calculate the effects of these disruptions. The details of our approach are contained below, after a presentation of the relevant literature.

Literature Review

Gaonkar and Viswanadham presented in 2004 a conceptual framework for the management of risk in supply chains (106). In it they described a classification of supply chain risk problems in 3 levels: the nature of the risk, the type of problem, and the planning level affected. According to its nature, our problem could be described as environmental; according with the type of problem, absenteeism could be categorized as a deviation, if it is mild, as a disruption if it would generate a total blockage of individual components of the network, and as a disaster if the SC would close (in this work we will always refer to them as disruptions); according with the planning level, the problem will usually affect the tactic level, although it could scale to the strategic level.

Food supply chain disruptions due to diseases are mentioned in the literature as examples of potential disruptions (106). A specific framework for the design of robust food supply chains was developed by Vljajic et al. (107), where characteristics associated with food supply chains, such as seasonality of supply, seasonality of demand, and

limited shelf-life, are presented as additional sources of vulnerability. They characterize disruptions (or disturbances) according to its size, as minor, major, or a failure.

Influenza risk analysis (108), explain that the expected severity of flu-like illnesses could range from negligible (localized asymptomatic infections) to disastrous (pandemic with severe syndromes or death). For our analysis we will assume pandemic infections comparable to the 2009 H1N1 influenza. Large scale disruptions to food supply chains due to diseases in animals that are transmissible to humans are expected to have high economic impacts (109). But compared to that scenario, pandemic illnesses could cause disruptions due to absenteeism caused by generalized infections, which could affect simultaneously several food supply chains, potentially affecting the final delivery of all foods to the general population.

BWE is first defined by Lee et.al (110) as a phenomenon where the variation of the orders placed to the supplier have a higher variation than the orders received from the customers. They also found that this effect is propagated and amplified upstream in the supply chain. Reverse BWE was defined by Rong et al. (111) as a similar effect propagating and being amplified downstream. Disruptions, understood either as a mismatch of forecast and demand at the lower end (final customers) of the supply chain (112), or as insufficient supply at the top (production) of the supply chain (113) have been related to the bullwhip effect (BWE) and the reverse bullwhip effect (RBWE), respectively. Although neither of these two causes directly relate to our problem, and we have not found literature relating disturbances to the internal connections of the supply chain to either of these effects, our preliminary results show bullwhip-like effects in our disrupted network. Also, disturbances to the flow of goods in the supply chain will

potentially generate supply and demand changes at each node of the network, which could generate a series of small BWE and RBWE (usually defined for sequential supply chains) along the entire network. For this reason, we review the BWE and RBWE literature, finding similarities and contrasts with other published materials.

Behavior of supply chain actors, from producers to consumers, appears invariably among the causes of bullwhip in the literature, although not always with the same interpretation. One of the first accounts of the phenomenon was presented in 1958 by Forrester (114) blaming it on the “irrational behavior of actors in an SC” (115), while a Lee, Padmanabhan, and Seungjin (110) discovered that BWE occurs “despite the rational behavior of all the actors in the SC” (115). Our preliminary setup for the problem proposes “mechanical” rules for the flow of goods, expressed in the form of inventory level policies, ordering and allocation strategies, and redundant capacities, not attached to any behavioral rules for the participants. Still, we observe bullwhip-like effects.

Rong, Snyder and Shen (111) concluded in 2008 that the sum of upstream and downstream bullwhip effects generated a larger variability in the orders at the center of the supply chain, and lesser at its extremes, forming what they called an “umbrella pattern”. They also presented in 2009 (113) that disruptions in the supply can cause price changes when there is a perceived shortage, which can in turn affect the demand. In our analysis we propose a different scenario, in which both ends of our supply chain are fixed through constant supply into the network and constant demand from the final consumers. The only disturbances affecting the behavior of the supply chain are the generalized disruptions caused by flu-induced absenteeism. What we observed in the preliminary results is that larger effects caused by the disruptions are observed in the middle echelons

of the supply chain, similarly to the mentioned bullwhip accounts. Following their sting metaphor that visualizes BWE and RBWE as small variations in one of the ends of the string being amplified towards the other end, our problem would show a string with both ends tied that grows variation from disturbances happening at places in the middle, more like a musical string or a vibrating trampoline than a bullwhip.

Lee et al. suggest four main causes for the BWE, none of them directly relates to our problem (110). Pricing (113, 116) does not affect our initial approach either. Many other causes of BWE focus mainly on responses of the supply chain to a variable demand (117-119), which does not relate to our initial approach that considers a constant demand, but could relate to the sub-supply chains formed by subsets of nodes in our network. Disney and Towill (120) refer to the commonly used order up to level (OUTL) inventory strategy as a major cause of BWE, while Csik and Foldesi (121) show that a safety stock proportional to the demand will cause instability similar to BWE in sequential supply chains. For our first approach, we chose a simplistic inventory policy, consisting of a fix inventory target proportional to the long run demand. This could be a cause for the bullwhip-type effects we observe in our preliminary results.

Collaboration and information sharing in the supply chain are shown to dampen BWE (122-124). Order smoothing may work as a remedy as well (124). Our initial approach is a basic representation of the supply chain and does not consider these more advanced features yet.

The great majority of modeling analyses on the bullwhip effect are done in sequential supply chains, which only have one player at each echelon of the chain. One study analyzes BWE in supply chain networks, in which members at different echelons of

the supply chain may interconnect with several other members, using frequency domain, a popular tool in the analysis of electronic circuits (125). One of the limitations of that analysis is that it does not consider capacities. Our proposed network is capacitated, and one of our decision parameters relates to capacity.

We will finally mention that seasonal and pandemic influenza cause absenteeism to the labor force; and pandemic influenza has been shown to cause higher absenteeism than seasonal in other countries (126, 127). Both have a negative impact on the economy, and this impact could be aggravated if policies such as school closures are enforced (128, 129). Our initial model only considers absenteeism in work groups caused by pandemic flu as simulated by Shi et al. (130), but does not consider absenteeism caused by policies.

Methods

Modeling the milk supply chain

For the sake of isolating the effects due solely to generalized disruptions, we assume no seasonality for production and consumption, and no other fluctuations in demand as a first approach. Our model assumes that the same amount of milk enters the supply chain every period, and serves the same constant demand at each of the end nodes of the supply chain every period. Given that less than 1% of the milk consumed in the U. S. is imported, and that less than 28% of the total milk is actually used as drinkable milk (131), a continuous and stable flow of raw material into the supply chain supply is not far from reality.

For simplicity, we will assume that there is only one retailer at the center of each county in the U. S., which will supply all of its inhabitants with drinkable milk. We also assume that all milk drinkers will purchase the milk from the county retailer, at the end of

the supply chain, and not in any other point. We found detailed information for milk production at state level, but not always at county level. Whenever we have the counties where the milk is produced, we will use that information. We will allow the rest of the raw production in the model to gravitate close to the geographic center of their state.

Although we are conscious that different industrial arrays may exist for producing milk (132), we will consider that all milk produced goes through 4 stages: Raw production, consolidation and conformation (e.g. pasteurization, homogenization, etc.), packaging, and retail. Whenever we found a processing center that encompasses more than one of these stages, we separate it into 2 independent processes with the same geographic location.

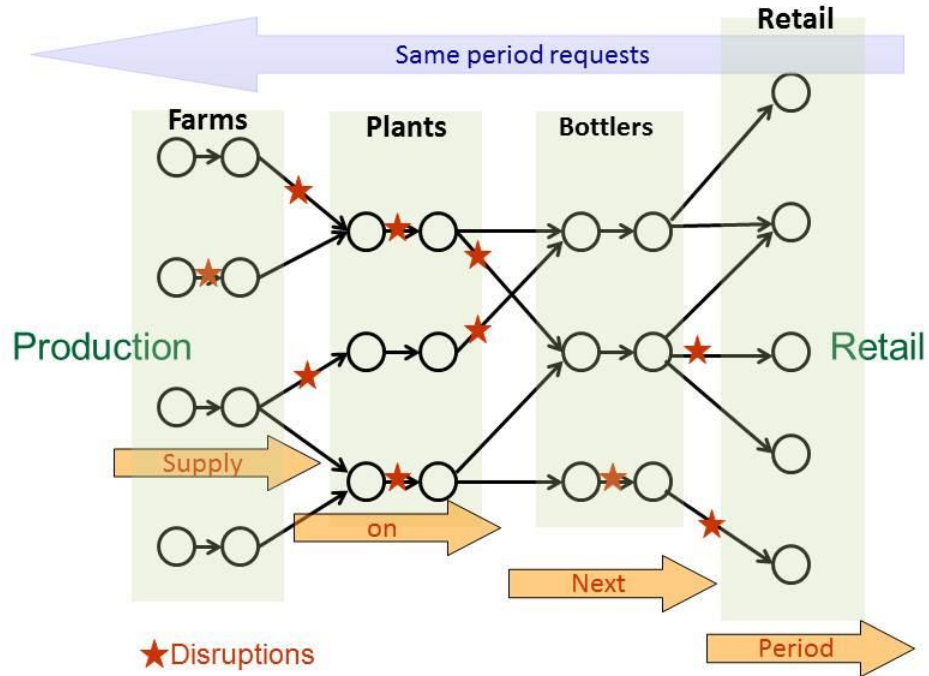
To construct the supply chain we found the largest possible amount of information related with the actual network, and the location of the different processing sites. When connections between sites are explicit, they will be fixed as an existing edge by adding a lower bound to the edge. All unknown connections require the generation of all feasible interactions among the supply chain nodes. For example, if a raw producer is known to sell all of its production to a single consolidator, then that link is settled and there will not be a need to further connect that raw producer to other potential processing plants. But if knowledge about the destination of the raw milk of this producer is unknown, then we generate all possible connections between that producer and all feasible processing plants in the next echelon. Feasibility rules such as the maximum length between nodes are used. We use optimization to propose minimum cost connections that could plausibly fill in for the unknown interactions. We assume that each producer will likely prefer to deliver their product to the closest possible location of

the next echelon. In our case, since we are using days as the shipping and processing period, we will not allow the milk to move more than 650 miles between echelons, which is approximately equivalent to the maximum of 11 hours that by law a driver is allowed to run per day.

The minimum cost approximation is adopted as the working supply chain. From it we will build a capacitated network. We assume that each of the three main manufacturing processes (all except retail) have a limited flow capacity that requires manpower to be operated, and that absenteeism will likely affect its operating capacity. To better model the capacity in these production nodes we split them in two parts. The link between them will represent the production flow inside these locations. Now, a capacity can be defined for each of the edges in the network.

After splitting the production nodes, our three production echelons (raw production, consolidation and conformation, and packing) become six modeled echelons, two nodes for each real echelon. Adding retail, we obtain 7 simulated echelons. A complete graphical description of the milk supply chain network model is shown in Figure 5.1. The first sets of two nodes from left to right represent the raw milk production farms. The next two nodes represent the consolidation and conformation plants, followed by the two nodes representing the bottling and delivering functions. The far right nodes represent retail locations. Milk flows from left to right and information from right to left. Stars are used to exemplify partial disruptions occurring on edges. Table 5.1 presents the number of nodes at each echelon, which in total add to 4217. The total number of active edges in the supply chain adds to 4411.

The nominal capacity of the arcs is defined by their flow in the absence of disruptions, plus an additional slack percentage. We will discuss more about this slack capacity in the sections below.



1

Figure 5.1: Graphic representation of the milk supply chain network and its first approach modeling components

Table 5.1: Number of nodes at each echelon of the milk supply chain.

Echelon	Description	Number of nodes
0	Raw production IN	132
1	Raw production OUT	132
2	Consolidation and Conf. Plants IN	134
3	Consolidation and Conf. Plants OUT	134
4	Packing IN	288
5	Packing OUT	288
6	Retail	3109
Total		4217

Timing for information flow and product supply

We assume that information will travel upstream, from one echelon to the next, in one period of time (for milk the period is assumed to be one day). Orders will be immediately supplied, but each of the downstream movements of product to the following echelon is assumed to last one full period. Purchase orders generated at each period will be received by the supplier in the same period, and the ordered product will be received in the next period.

Ordering, inventory and service policies

Our supply chain is assumed to be decentralized. Therefore, each of its parts will act independently of the others, according with its own established policies. This assumption poses a worst case scenario, given that collaborative supply chains can usually improve their reaction to variations (122-124). We believe this assumption may also be a closer representation to absenteeism due to a flu-like illness, which occurs randomly and suddenly and last a brief but random number of periods. These characteristics may avert the benefits of collaboration.

A fixed inventory target level (a constant quantity to refill the stock up to) is adopted as the base approach to an inventory policy of all the nodes. Fixed target levels can result from having a constrained storage capacity or a defined maximum inventory level. Initially, we will fix the inventory target value proportional to the long term demand at each node.

When there is more than one supplier, ordered quantities are split always in the same proportion, according to the long term demand. When more than one buyer, supplies meet last received order, except in the case when there is no sufficient inventory

to supply both requests. In that case, supply is shared proportionally to the requested quantity. In this case, no priority is established for the orders of the customers. Since we are considering no seasonal effects, long term averages are considered equal to those quantities requested in the absence of disruptions and under equilibrium (at each node, flow in equals flow out, and all inventory levels reach their target value).

Calculation of order quantities

We assume supply chain events occur in the following sequence: 1) Previous period orders are received; 2) inventory levels are calculated; 3) orders from echelons down the supply chain are received; and 4) new orders are calculated and sent to suppliers.

The calculation of the sum of orders P_j^t to be placed by node j in period t is as follows. Let R_j^t be the sum of shipments received in period t by node j ; let I_j^t be the inventory level calculated at period t , at node j ; let O_j^t be the sum of orders received by node j in period t ; and let P_j^t be the sum of orders placed by node j in period t . Let I_{max} represent the target inventory. Then: $P_j^t = \min\{I_{max} - I_j^t - R_j^t + O_j^t, I_{max} - R_j^t\}$.

Disrupting the supply chain

We use absenteeism groups generated by an agent based model that simulates possible infections due to interactions of individuals that belong to different groups (133). Using these groups' absenteeism outcomes from that model (ran for $R_0 = 1.3$) allows us to randomly relate work groups in the Supply Chain with the simulated group's absenteeism outcome. We randomly assign different absenteeism behaviors to each edge. This will give a particular timing and acuteness of the disease, to each of the different work groups

in the milk's Supply Chain. Disruptions are caused by absenteeism in all milk processing sites (edges between duplicated nodes) and in their associated transportation service, this is, the working groups that transport goods between different processes (edges between a node out of one process and a node into the next process or retail). Disruptions are generalized along the entire network.

As a first approach, no disruptions are considered for the final consumer or at the raw milk source. The assumption that there are no disruptions at the end consumer considers that illness will not affect the consumption behavior for this particular food item. The assumption that there will be no disruptions affecting the milk's input source acknowledges a much larger production of raw milk than that used for drinkable milk and a source (cows) unaffected by the pandemic.

Simulation

We constructed a Monte Carlo simulation program to randomly apply disruptions to the different arcs of the supply network using C++. Initially, we considered 10% to be a normal slack on transportation and processing capacities, which will enable the network to absorb small changes in the supply patterns. We will also consider an inventory level of two average periods, in this case two days of average demand, as the order-up-to quantity, at each of the processing sites. We simulate 1000 replications under these conditions, randomly assigning a different absenteeism group at each replication. Each replication simulates absenteeism for 180 days.

Refining the simulation capabilities for a more elaborated approach, we changed the percentage of additional capacity available between 0 and 20%, on increments of 1% each time, keeping the order-up-to quantity fixed at 2 days. The simulation ran 1000

replications for each of these scenarios. We also allowed the simulation to keep the slack capacity at 10% while varying the order target inventory level between 1 and 3 days, where 1 day represents the “just in time” quantity (assuming one shipment a day).

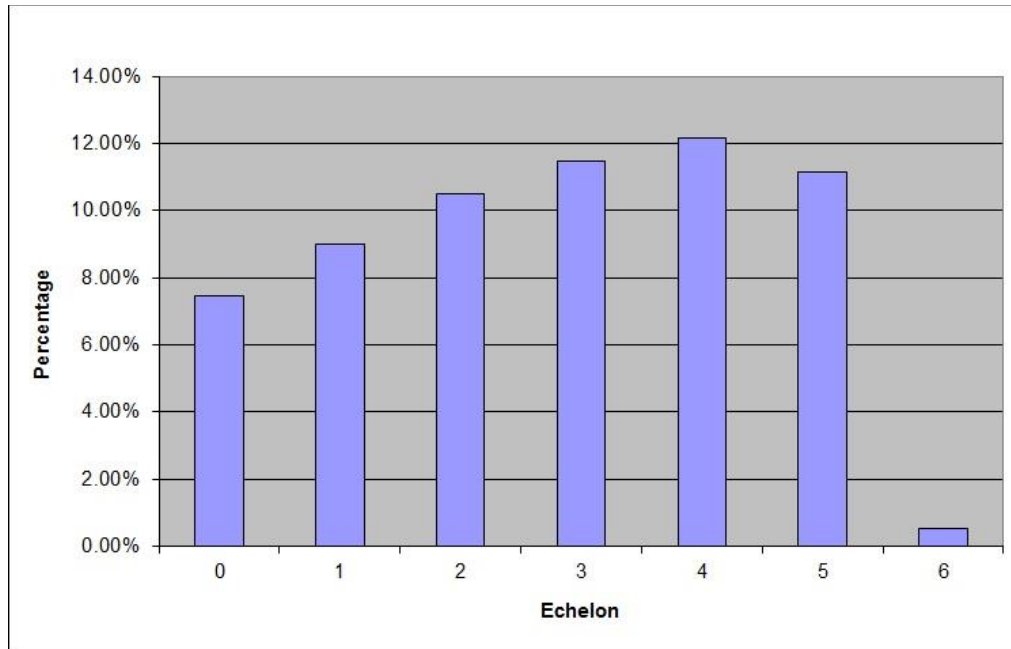


Figure 5.2: Percentage of periods with unmet demand with 10% of slack capacity and 2 days of target inventory

Preliminary Results

Fixed slack capacity and target inventory levels

We build a base case with 10% of slack capacity and 2 days of target inventory. After running a simulation for this basic case, we obtain the graph shown on Figure 5.2. Echelon 0 represents the cows, and echelon 6 represents retail. Notice that retailers (node 6) experience the shortest average unmet demand to its customers. Changes in the

demanded quantities set by the independent decision makers in the different nodes of the network induce variability along the Supply Chain. Here, the demand to the retailer is assumed constant. Also, in the same figure we can appreciate that the largest percentage of unmet demand is found at the intermediate processing stages.

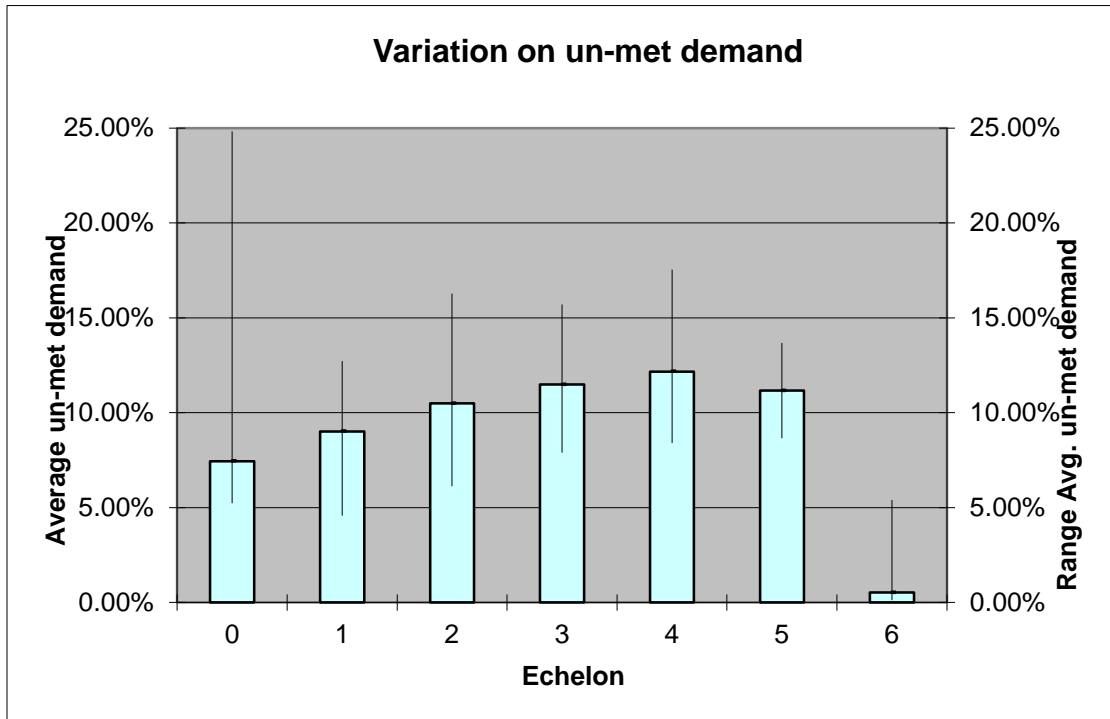


Figure 5.3: Variation on the percentage of periods with unmet demand when slack capacity is 10% and the target inventory is 2 days

The demand variability provoked by the disruptions is mainly experienced at the zeroth node. Figure 5.3 shows this increase in variability.

The average accumulation of inventory (measured in days of inventory) at each echelon level reduces as the supply chain approaches the final consumer (see Figure 5.4). This can be explained by the accumulation of the impact of capacity reduction moving

downstream in the supply chain, which increasingly restricts the flow of goods from one level to the next.

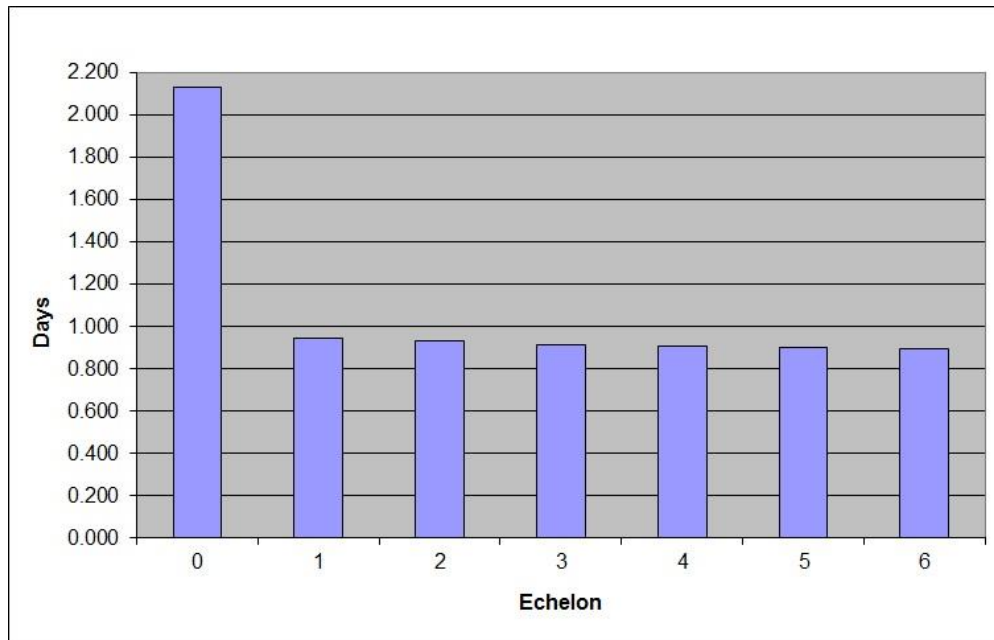


Figure 5.4: Average days of inventory accumulated at each echelon when slack capacity is 10% and the target inventory is 2 days

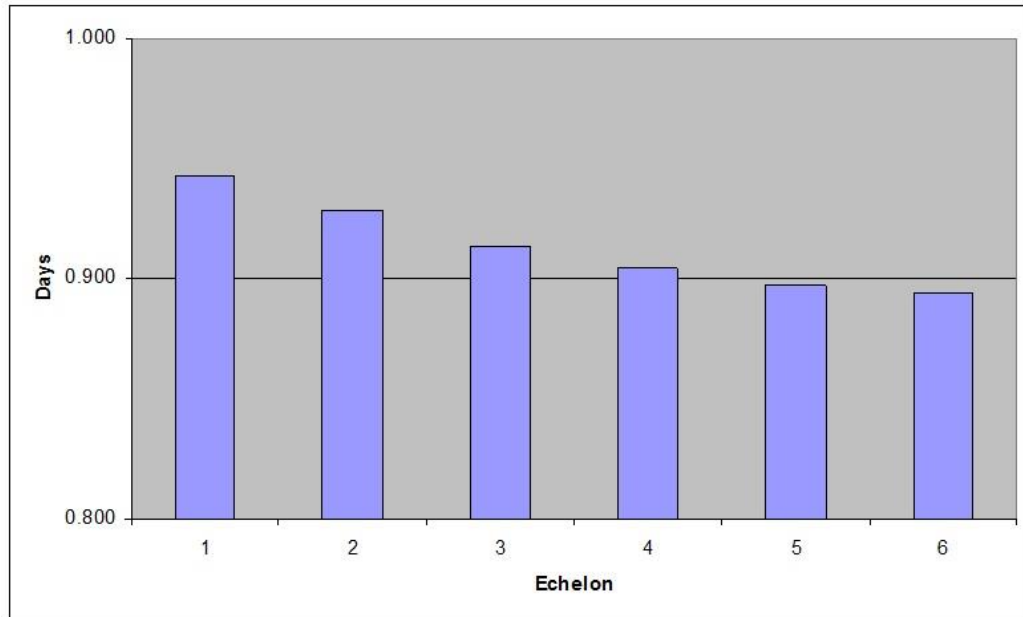


Figure 5.5: Average days of inventory accumulated at echelons 1 - 6 when additional capacity is 10% and the target inventory is 2 days

On Figure 5.5 we focus our attention on echelons 1 through 6 (removing echelon 0 from the graph). Notice that the change in the inventory level (measured in days of inventory) decreases monotonically.

Finally, fixing both slack capacity and target inventory we calculate the average percentage of times that inventory level reaches zero at each echelon. Surprisingly we observe that the effect alternates from one node level to the next. Observing Figure 5.6 we also appreciate that the greatest oscillations occur in one of the intermediate echelon levels. We notice that echelons 1, 3 and 5 have a single connection with echelons 0, 2, and 4, which are the input nodes to the 3 processes. Contrarily, each node in echelon 3 connects in average with 2 nodes in echelon 4, and each node in echelon 5 connects in average with 11 nodes in echelon 6. The difference between the variance of a single connection as compared to the variance with multiple connections could cause this effect.

Notice that even though retailers in node 6 have a significant number of hits to zero inventory level, the difference between the supply and the constant demand does not add up significantly in the amount of unmet demand, as happens with any other node subject to larger differences between supply and demand (Figure 5.2).

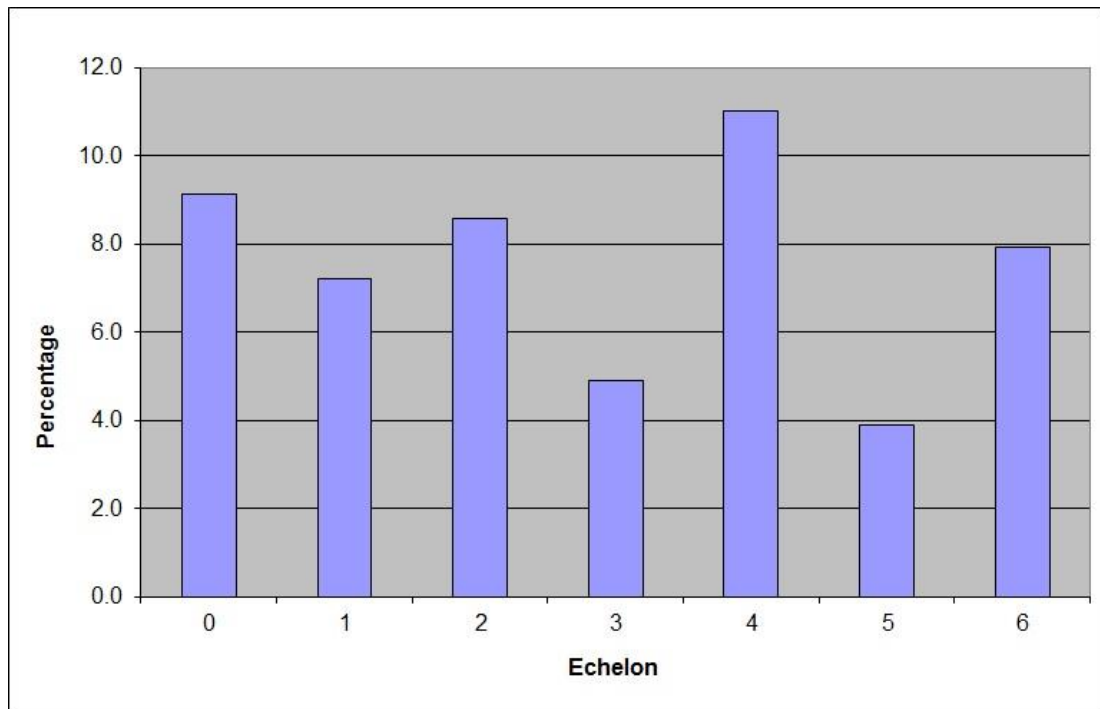


Figure 5.6: Average percentage of periods that reach an inventory level of zero when additional capacity is 10% and the target inventory is 2 days.

Variable slack capacity level with fixed target inventory level

When target inventory level is kept constant at 2 days of inventory, an increase in slack capacity (additional percentage of available capacity) causes the average inventory level to decrease in the first echelons. This is more notorious at echelon 0. Figure 5.7 shows that when the additional capacity is reduced drastically, the inventory kept at the

initial echelon grows (because the restricted supply chain will not stop the constant supply on node 0), and inventory on echelon 6 also reduces.

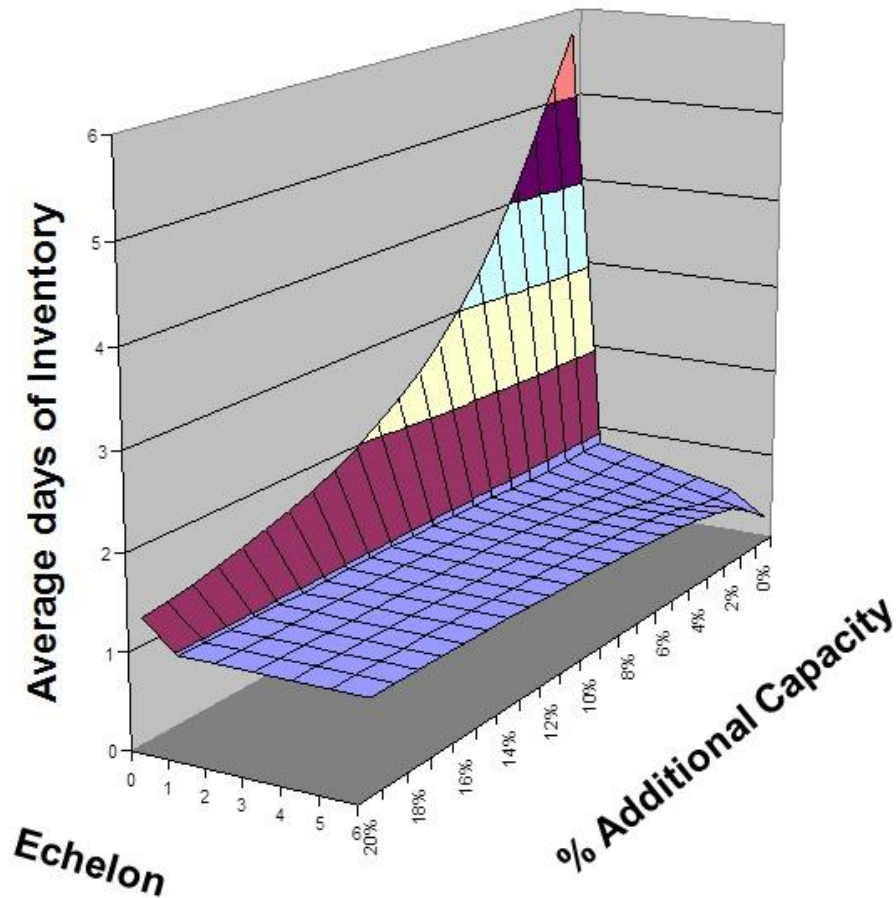


Figure 5.7: Average days of inventory at different levels of slack capacity and target inventory of 2 days.

As additional capacity is decreased, the proportion of unmet demand grows for all echelons. But the growth in the unmet demand for echelons 4 and 5 is very steep, reaching around 70% for echelon 5 (see Figure 5.8). Surprisingly, the growth in the

proportion of unmet demand for echelon 6 is barely noticeable. Apparently, the unmet demand in the last echelon depends more strongly on the ordering-up-to policy (amount of safety stock), and not so much by the restricted flow that follows disruptions.

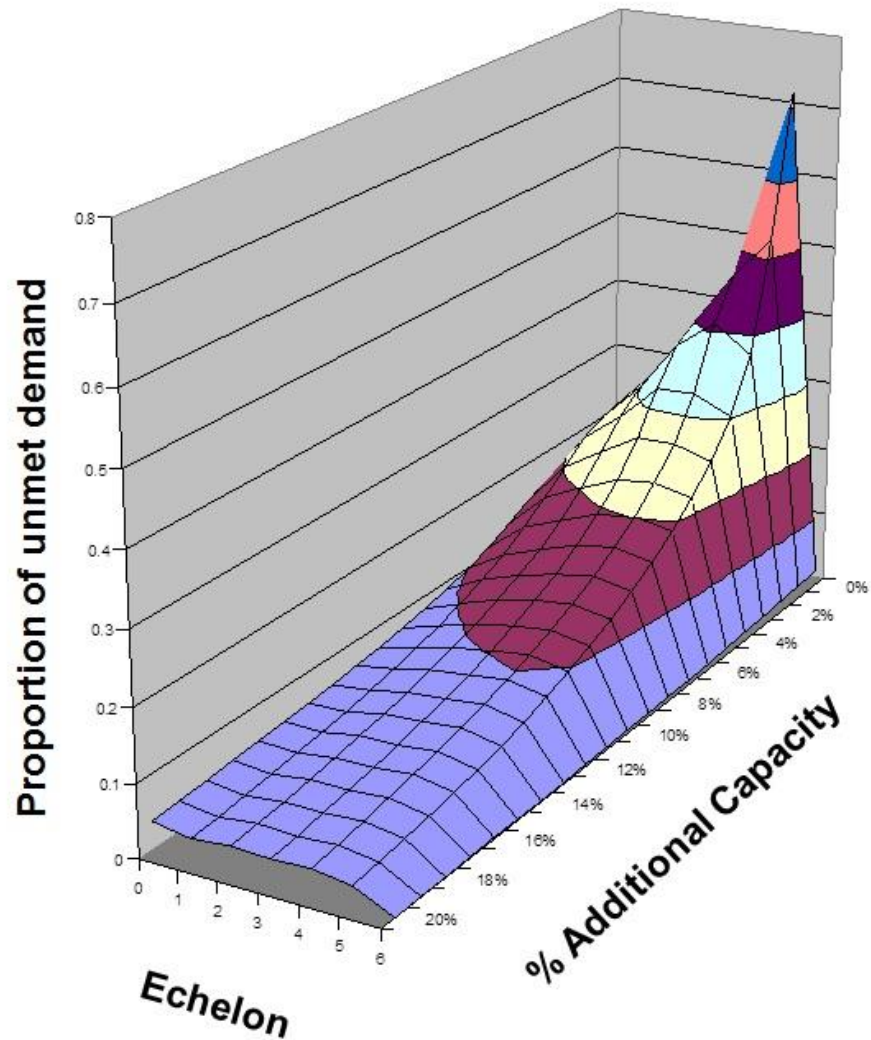


Figure 5.8: Average percentage of unmet demand per echelon at different levels of slack capacity and target inventory of 2 days.

The “trampoline effect” on the average number of times inventory drops to zero appears as the slack capacity drops below 15% and increases as the additional capacity is reduced (see Figure 5.9). As additional capacity is increased, the average number of times inventory hits 0 decreases for all echelons, except for the production node, which then has an increased flow of its inventory into the rest of the supply chain, and in turn renders more occurrences of zero inventory.

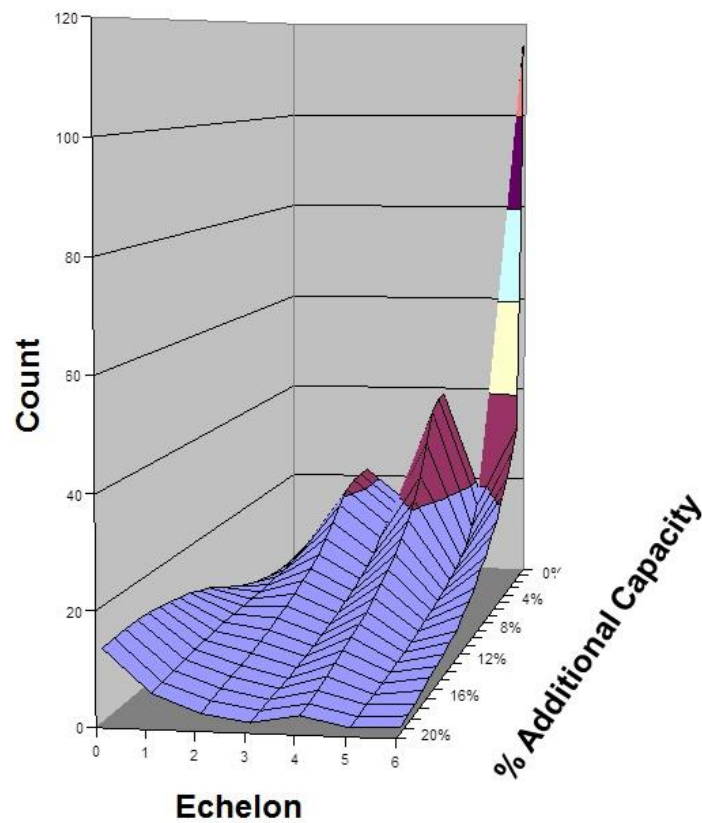


Figure 5.9: Average count of periods that inventory reaches a zero level at different levels of slack capacity and target inventory of 2 days.

Fixed slack capacity level with variable target inventory level

Now, we let the slack capacity of the system have a steady value of 10% while we move the target level of inventory between 1 and 3 days. When we do this, we find the realization of the intuitive fact that a greater amount of order-up-to days in the ordering policy will increase the average inventory; But not for all the echelons. Average inventory at echelon zero shows a convex behavior, with a minimum around 1.9 order-up-to days, as can be appreciated on Figure 5.10.

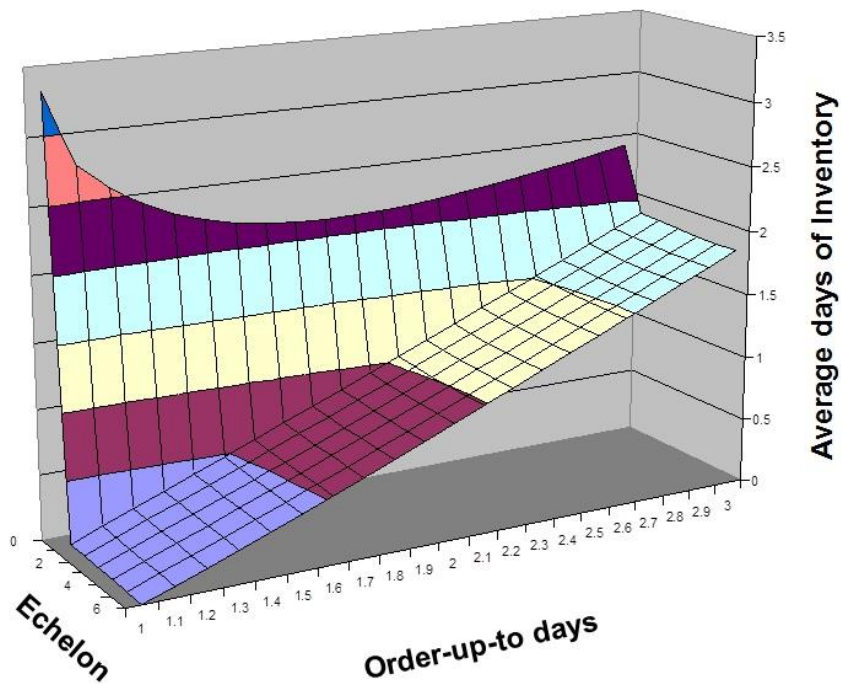


Figure 5.10: Average inventory level at a 10% slack capacity and variable target inventory.

The fact that the proportion of unmet demand increases as the order-up-to days increases for the other echelons seems counter-intuitive. Why is it that the bigger the

stock the nodes want to form, the higher is the proportion of the unmet demand? That is because a greater request for a good that is produced at a constant rate will more likely generate orders that cannot be fulfilled by upstream echelons. However, notice on Figure 5.11 that the proportion of unmet demand reduces for the final costumers, due to the pull effect that the entire supply chain is exerting towards the retailer.

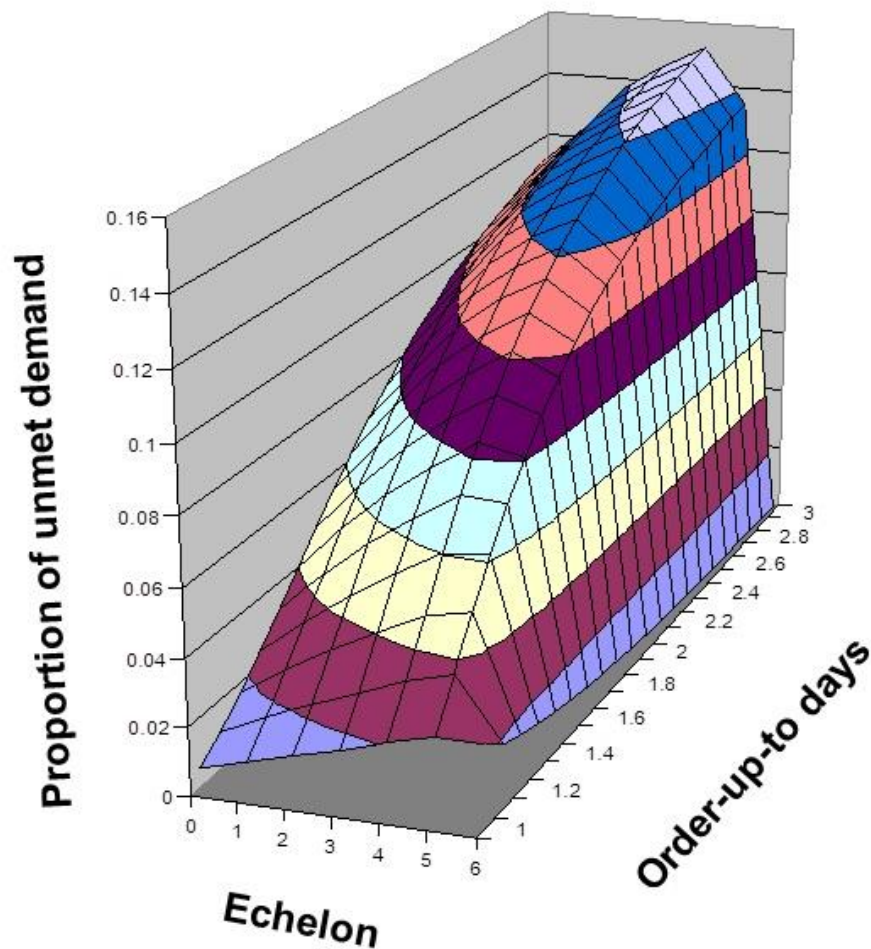


Figure 5.11: Average inventory level at a 10% slack capacity and variable target inventory.

Once more the alternate effect in the number of zeros in the inventory level can be observed for most of the order-up-to levels. The greater the target inventory is, the more marked is the effect. The behavior of echelon zero differs when the target inventory reduces, but every other echelon increases rapidly (and still alternately) their count of zero level inventory periods (see Figure 5.12).

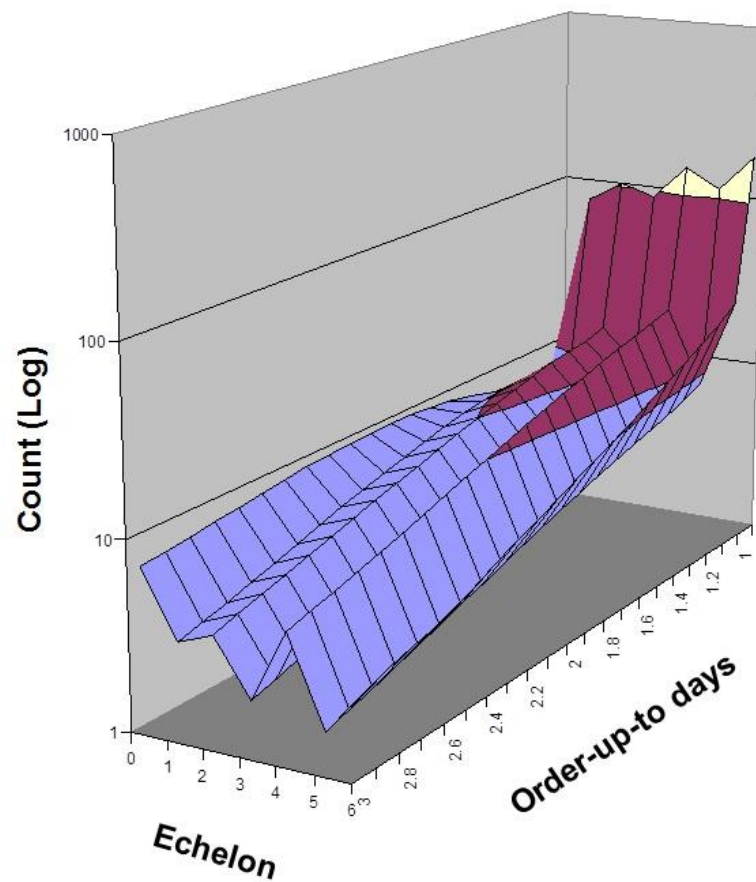


Figure 5.12: Average count of periods with zero inventory level at a 10% slack capacity and variable target inventory.

Discussion of preliminary results

One of the most characteristic findings in the preliminary results is the observed “trampoline effect”, in which the number of times inventory hits zero forms a zig-zag pattern between echelons. The model clearly shows the big impact that generalized absenteeism may have on the overall performance of the supply chain, even when no explicit changes are applied to the total input or output of the system. It also confirms how small variations to the equilibrium of supply chains can cause extensive modifications to its behavior by the multiplying effect carried by the individual reactions.

It is clear that this analysis can be applied to pandemic diseases of different severities and other similar disruptions from different causes, which gives added relevance to this study.

Different echelons experience different consequences from the disruptions, and these do not always have a monotonic intensity through the supply chain. These differences may depend on the network array, given the number of connections a node may have coming into it, compared to the number of connections going out of it.

Considering capacity by splitting the nodes of production facilities adds an interesting perspective of how internal connections in these plants play a relevant role in the overall analysis of the supply chain.

APPENDIX A

COMPREHENSIVE LIST OF ANALYZED VARIABLES

POTENTIALLY RELATED TO H1N1 VACCINATION COVERAGE

The table below contains the list of factors analyzed in chapters 2 and 3 of this thesis.

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
Group 1: Coverage measurements (Independent variables)								
y6	Coverage on Children 6 mos. to 17 yrs.	MMWR (7)	38.9	11.9	84.7	21.3		
y7	Coverage on persons aged ≥ 18 yrs.	MMWR (7)	19.9	5.3	34.4	8.7		
y8	Coverage on persons aged ≥ 6 mos.	MMWR (7)	24.5	6.1	38.8	12.9		
y9	Coverage on persons aged 25--64 yrs. at high risk	MMWR (7)	25.4	7.6	47.2	10.4		
Dependent variables								
Group 2: Population composition								
c0	Resident population 18 years and over, but under 65 years, percent (July 1 2009 - estimate)	Calculation from Census (17)	56.2	2.1	63.1	50.2		
c1	Resident population 65 years and over, percent (July 1 2009 - estimate)	Census (17)	13	1.7	17.4	7.3		
c2	Resident population 5 - 17 years, percent (July 1 2009 - estimate)	Census (17)	24	1.9	31	18.9		
c3	Resident population under 5 years, percent (July 1 2009 - estimate)	Census (17)	6.8	0.8	9.8	5.3		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
c4	Resident population: Black alone, percent (July 1 2009 - estimate)	Census (17)	11.4	11.1	54.4	0.7		
c5	Resident population: Not Hispanic, White alone, percent (July 1 2009 - estimate)	Census (17)	72.3	15.9	95.3	24.9		
c6	Resident population: White alone, percent (July 1 2009 - estimate)	Census (17)	81.1	13.1	96.4	29.7		
c7	Resident population: Hispanic or Latino Origin, percent (July 1 2009 - estimate)	Census (17)	9.8	9.6	44.9	1.1		
c8	Percent foreign born population 2000 (sample)	Census (18)	7.3	5.6	26.2	1.1		
c9	Population 5 years and over, percent speaking language other than English at home 2000 (sample)	Census (18)	12.7	8.8	39.5	2.7		
c19	Resident population: American Indian and Alaska Native alone, percent (July 1 2009 - estimate)	Census (17)	1.8	2.9	15.3	0.2		
co4	Population under 10 years old, 2009.	Census (17)	13.5	1.4	18.78	10.66		
Group 3: Population economic and other social characteristics								
c10	Educational attainment - persons 25 years and over - percent high school graduate or higher	Census (18)	81.9	4.3	88.3	72.9		
c11	Educational attainment - persons 25 years and over - percent bachelor's degree or higher	Census (18)	24.1	4.7	39.1	14.8		
c12	Average travel time to work for workers 16 years and over not working at home	Census (18)	23.7	3.5	31.7	15.8		
c13	Per capita personal income 2007	Census (18)	37825	6594	62484	28541		
c14	Median Family Income in 2008	Census (18)	63715	9583	85761	46668		
c15	Two year average median household income 2007-2008	Census (18)	51557	7527	68175	37579		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
c16	People of all ages in poverty - percent 2007	Census (18)	12.7	3	20.7	7.3		
co1	Calculation: Births 2006 * 100 / Population 2005	Census from Census (18)	1.4	0.2	2.1	1.1		
co2	Persons per household 2000 (complete count)	Census (18)	2.6	0.1	3.13	2.16		
co3	Percentage of children living with their grandparents	Census (134)	5.2	2.9	18.9	1.5		
co5	Total number of separated, widowed and divorced people	Census (135)	0.13	0.02	0.17	0.08		
ii1	Income inequalities measured through the mean logarithmic deviation method	Volscho (2009) (53)	0.36	0.05	0.58	0.29		
ii2	Income inequalities measured by the Theil's entropy index	Volscho (2009) (53)	0.34	0.04	0.51	0.27		
ii3	Income inequalities measured as 1/2 of the squared coefficient of variation	Volscho (2009) (53)	0.51	0.08	0.8	0.33		
ii4	Income inequalities measured by the Gini index	Volscho (2009) (53)	0.43	0.03	0.53	0.39		
ii5	Income inequalities measured by Atkinson's Index with inequality aversion parameter $e = 0.5$	Volscho (2009) (53)	0.16	0.02	0.24	0.13		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
ii6	Income inequalities measured by Atkinson's Index with inequality aversion parameter $e = 1.0$	Volscho (2009) (53)	0.3	0.03	0.44	0.26		
ii7	Income inequalities measured by Atkinson's Index with inequality aversion parameter $e = 2.0$	Volscho (2009) (53)	0.75	0.05	0.85	0.61		
ns1	Neighborhood Segregation: Index of Dissimilation for all minorities - white	James et al (2009) (54)	0.3	0.1	0.75	0.08		
ns2	Neighborhood Segregation: Index of Dissimilation for black - white	James et al (2009) (54)	0.4	0.1	0.82	0.1		
ns3	Neighborhood Segregation: Index of Dissimilation for Hispanic - white	James et al (2009) (54)	0.3	0.1	0.6	0.04		
ns4	Neighborhood Segregation: Index of Dissimilation for Asian and NHPI - white	James et al (2009) (54)	0.3	0.1	0.49	0.11		
whd3	Social Determinants Dimension Scores	James et al (2009) (54)	0	0.64	1.37	-1.73		
Group 4: State general characteristics								
pop	Estimated total State population by July 1, 2009	Census (17)	6019736	6713840	36961664	544270		
c18	Population per square mile 2000	Census (17)	362	1298	9378	1.1		
area	Total area of the state	Census (17)	74394	95725	663267	68		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
area_sqrt	Square root of area	Calculation from Census (17)	241.5	126.7	814.4	8.3		
lat	Average latitude	Census (17)	39.5	6	61.4	21.1		
long	Average longitude	Census (17)	-93.3	19.1	-69.4	-157.5		
nc	Number of Counties in the state	Census (17)	61.6	46.4	254	1		
Group 5: State economics and related policy								
c17	Federal Government expenditure per capita FY 2008	Census (18)	10680	9947	79757	6255		
hr	1- Home Rule Local, 2- Home Rule Mixed, 3- Home Rule State; Who holds governance of local health departments	2008 National profile of Local Health Departments (22)	2.5	0.7	3	1		
hrl	1 if Localities hold governance of local health departments, 0 if not.	2008 National profile of Local Health Departments (22)	0.1	0.3	1	0	6	45
hrm	1 if governance of local health departments is shared between localities and state	2008 National profile of Local Health Departments (22)	0.3	0.4	1	0	13	38
hrs	1 if State holds governance of local health departments, 0 if not.	2008 National profile of Local Health Departments (22)	0.6	0.5	1	0	32	19
t1	Total number of cars, trucks and buses per capita	National Auto Dealers Association (21)	0.8	0.2	1.19	0.36		
t2	Number of driving licenses divided by population	StateMaster.com (136)	68.3	5.5	87.2	56.9		
tb	1- State uses Thimerosal 0- State Bans Thimerosal	NewsMax.com (137)	0.86	0.34	1	0	44	7

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
Group 6: Health infrastructure								
ah14	Public Health Funding. State funding dedicated to public health as well as federal funding directed to states by the Centers for Disease Control and Prevention and the Health Resources and Services Administration, expressed on a per capita basis.	America's Health Rankings (28)	81.2	39.8	220.4	35.4		
ah22	Medicaid expenditure in Long Term care (calculated as total Medicaid spending times percentage spent on long term care) in millions of dollars	State Health Facts (25)	2251	3274	20333	236		
ah23	Medicaid expenditure in Long Term care per capita	Calculation from State Health Facts (25)	384	156	1041	143		
ah24	Medicare enrollment as percent of total population	State Health Facts (25)	15.5	2.2	21	9		
ah25	Percentage Reporting Not Seeing a Doctor in the Past 12 Months Because of Cost	State Health Facts (25)	13	3.4	20.5	6.2		
ah3	Have any kind of health care coverage (percentage of adults)	BRFSS (26)	85.6	4.2	94.7	74.8		
ah28	Number of Residents in Certified Nursing Facilities, 2008	State Health Facts (25)	27223	26983	109257	605		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
ah29	Estimated Underserved Population Living in Primary Care Health Professional Shortage Areas (HPSAs), as of September, 2008	State Health Facts (25)	12.6	7.6	34.4	1.7		
ah30	Certified Nursing Facility Beds, 2008	State Health Facts (25)	32326	31862	123367	698		
ah34	Health Care Expenditures by State of Residence (in millions), 2004	State Health Facts (25)	30417	32873	166236	2662		
ah35	Health Care Expenditures by State of Residence per capita, 2004	Calculation from State Health Facts (25)	5191	830	8020	3454		
ah39	Number of Federally-Funded Federally Qualified Health Centers, 2008	State Health Facts (25)	20.6	17.3	113	2		
ah40	Percentage of non-federal physicians that are Non-Hispanic White, 2008	State Health Facts (25)	48.3	7.5	64	29		
ch3	Child has any kind of health insurance	National Survey of Children's Health (56)	91.6	3.5	97.3	80.6		
ch4	Child has a private insurance	National Survey of Children's Health (56)	63	7.8	76	46.6		
pr1	Number of Health Care Practitioners 2008	Bureau of Labor Statistics (23)	24.45	4.38	43.27	16.75		
pr2	Active Physicians per thousand population, 2006	American Medical Association (24)	2.68	0.97	7.99	1.69		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
whd2	Access and Utilization Dimension Scores	James et al (2009) (54)	0	0.74	1.58	-1.3		
whd4	Physician diversity ratio	James et al (2009) (54)	3.82	2.09	11.53	0.91		
whd5	Percent of women living in a Primary Care Health Professional	James et al (2009) (54)	0.44	0.08	0.61	0.22		
Group 7: Health conditions, adults								
ah1	Adults who have been told they currently have asthma	BRFSS (26)	8.7	1.2	11.1	6.3		
ah2	Adults who have been told by a doctor that they have diabetes (perc.)	BRFSS (26)	8.5	1.6	12.3	5.7		
ah4	Good or better health reported as health status	BRFSS (26)	85	3.5	89.8	76.2		
ah5	Weight classification by BMI as overweight [25-30) or obese [30+)	BRFSS (26)	63.4	3.6	70.2	51.7		
ah6	Cumulative adolescent or adult AIDS cases reported	CDC (138)	17969	32859	170035	139		
ah9	Cancer Deaths per 100000 Population	America's Health Rankings (28)	194	16.3	225.1	144.7		
ah10	Cardiovascular Deaths per 100000 Population	America's Health Rankings (28)	283.6	38.7	378.5	212.6		
ah11	Percentage of population ever diagnosed with angina or coronary heart disease	America's Health Rankings (28)	4.3	0.9	8.1	2.5		
ah12	Percentage of adult population that has ever suffered a heart attack	America's Health Rankings (28)	4.3	0.9	7.6	2.2		
ah13	Preventable hospitalization enrollments per 1000 enrollees. Discharge rate among the Medicare population for diagnoses that are amenable to non-hospital based care.	America's Health Rankings (28)	71.4	15.9	109.3	29.3		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
ah18	Invasive Cancer Incidence Rate per 100,000 Population, 2007	State Health Facts (25)	464.4	32.5	536.1	394.1		
ah19	Number of adults with serious mental illness	State Health Facts (25)	228149	251165	1385837	20323		
ah20	Number of adults with serious mental illness per 100000 population	State Health (25)	3821	245	4258	3109		
ah21	Number of Deaths Caused by Stroke and other Cerebrovascular Diseases per 100,000 Population, 2007	State Health Facts (25)	42.7	6.4	57.4	28.2		
ah26	Hospital Emergency Room Visits per 1,000 Population, 2008	State Health (25)	429.2	92.1	740	275		
ah27	Hospital Outpatient Visits per 1,000 Population, 2008	State Health Facts (25)	2316	826	5323	1115		
ah31	Retail Prescription Drugs Filled at Pharmacies (Annual per Capita), 2009	State Health (25)	12.5	2.8	18.9	6.4		
ah32	Total Retail Sales (in millions of dollars) for Prescription Drugs Filled at Pharmacies, 2009	State Health Facts (25)	4258	4311	18908	350		
ah33	Total Retail Sales for Prescription Drugs Filled at Pharmacies, 2009	State Health Facts (25)	730.9	146.9	1079.3	366.2		
ah41	Chronic Liver Disease Mortality (2005)	CDC Chronic disease indicators (139)	8.9	1.9	14.7	5.9		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
ah42	Mortality with chronic obstructive pulmonary disease in adults 45+ (2005)	CDC Chronic disease indicators (139)	261.1	50.2	411.2	133.1		
ah43	Mortality with end-stage renal disease (2007)	CDC Chronic disease indicators (139)	26	7.7	52.5	13.8		
ah44	Incidence of treated end-stage renal disease (2006)	CDC Chronic disease indicators (139)	341.6	55.7	451.3	183.8		
whd1	Health Status Dimension Scores	James et al (2009) (54)	0	0.52	1.5	-0.85		
Group 8: Health conditions, children								
ch1	General health condition of child described as fair/poor	National Survey of Children's Health (56)	2.8	1	6.2	1.2		
ch2	Child has one or more chronic health condition	National Survey of Children's Health (56)	23	3	27.9	17.4		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
Group 9: Immunization coverage other than H1N1								
ah15	The average of the percentage of children ages 19 to 35 months who have received four or more doses of DTP, three or more doses of poliovirus vaccine, one or more doses of any measles-containing vaccine and three or more doses of HepB vaccine. This does not measure if each child received all series.	America's Health Rankings (28)	77	4.3	85	65.5		
ah17	Percentage of adults aged 65+ who have ever had a pneumonia vaccine.	State Health Facts (25)	68.1	3	73.9	59.8		
ov1	Children aged 19--35 mos. with ≥4 doses of DTaP, ≥3 doses of poliovirus vaccine, ≥1 dose of any measles-containing vaccine, ≥3 doses of Hib vaccine, ≥3 doses of hepatitis B vaccine, ≥1 dose of varicella vaccine in 2008	MMWR (140)	74.4	5.1	82.3	59.2		
ov2	Teenagers with ≥2 doses of measles, mumps, and rubella vaccine in 2008	MMWR (141)	89.2	5.3	99.5	73.2		
ov3	Teenagers with ≥1 doses of tetanus and diphtheria toxoids vaccine (Td), tetanus toxoid, reduced diphtheria toxoid, and acellular pertussis (Tdap), or tetanus-unknown type vaccine	MMWR (141)	70.1	12.9	94.4	28.7		
sf1	Seasonal Flu Coverage for high risk adults 18 -- 49 yrs. on the 2007-2008 season	CDC Influenza Vaccination Coverage (55)	32.8	7.5	50.8	14.3		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
sf1.2	Seasonal Flu Coverage for adults 18 -- 49 yrs. on the 2007-2008 season	CDC Influenza Vaccination Coverage (55)	55.4	11.5	80.5	27.3		
sf2	Seasonal Flu Coverage for non-high risk adults 18 -- 49 yrs. on the 2007-2008 season	CDC Influenza Vaccination Coverage (55)	22.6	5.1	37.8	11.9		
sf3	Seasonal Flu Coverage for adults 50 -- 64 yrs. on the 2007-2008 season	CDC Influenza Vaccination Coverage (55)	41.5	5.2	53.5	26.1		
sf4	Seasonal Flu Coverage for adults ≥ 65 yrs. on the 2007-2008 season	CDC Influenza Vaccination Coverage (55)	68.7	5	76.9	54.1		
sf5	Seasonal Flu Coverage for high risk adults 18 -- 49 yrs. on the 2006-2007 season	CDC Influenza Vaccination Coverage (55)	37.5	5.7	54	22.3		
sf6	Seasonal Flu Coverage for adults 18 -- 49 yrs. on the 2006-2007 season	CDC Influenza Vaccination Coverage (55)	25.1	3.8	35.9	17.5		
sf7	Seasonal Flu Coverage for adults 50 -- 64 yrs. on the 2006-2007 season	CDC Influenza Vaccination Coverage (55)	44.1	4.4	53.7	34		
sf8	Seasonal Flu Coverage for adults ≥ 65 yrs. on the 2006-2007 season	CDC Influenza Vaccination Coverage (55)	73.2	4.4	81	60.8		
sf9	Children 6 to 35 mos. that received full seasonal flu vaccination	MMWR (142)	23.6	7.3	40.7	7.1		
Group 10: Use of preventive health services (other than immunizations)								

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
ah7	Adults aged 50+ who have had a blood stool test within the past two years (2008)	BRFSS (26)	21.2	3.6	29	9.1		
ah8	Adults aged 50+ who have ever had a sigmoidoscopy or colonoscopy (2008)	BRFSS (26)	62.8	5.6	74.3	52.6		
ah16	Percent of adults who responded they had visited the dentist or dental clinic within the past year for any reason.	America's Health Rankings (28)	70.3	5.1	80.2	57.9		
ah36	Percent of Women Age 18 and Older Who Report Having Had a Pap Smear Within the Last Three Years, 2008	State Health Facts (25)	82.7	2.9	88.9	74.1		
ah37	Percent of Women Age 40 and Older Who Report Having Had a Mammogram Within the Last Two Years, 2008	State Health Facts (25)	75.8	4.5	84.8	67.1		
ah38	Percent of Women Age 50 or Older Who Report Ever Having Had a Colorectal Cancer Screening, 2008	State Health Facts (25)	62.4	5.6	73.2	51.8		
ch5	Child had one or more preventive medical care visits	National Survey of Children's Health (56)	87.8	5.1	97.7	76.7		
ch6	Child care meets the criteria for medical home	National Survey of Children's Health (56)	59.6	5.5	69.3	45.4		
Group 11: H1N1 Funding								
pf1	FY 2009 Funding Total based on 2009 Omnibus-enacted numbers (per capita)	CDC (29) (Calculation)	3.3	2.1	10.78	1.73		
pf2	Phase III Total Funds Awarded September 2009 (per capita)	CDC (30)	3.25	0.9	5.97	2.38		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
pf3	Phases I and II % Unobligated through January 31, 2010	CDC Report (29)	0.22	0.21	0.72	0		
pf4	Phase III % Unobligated through January 31, 2010	CDC Report (29)	0.28	0.26	0.95	0		
Group 12: H1N1 Epidemic								
pw1	Peak week for number of ILI cases after week 30	Calculation from ILINet (31)	41.6	2.2	46	35		
pw2	Peak number of cases after week 30, per thousand population	Calculation from ILINet (31)	0.41	0.36	1.69	0.06		
pw3	Percentage of weeks with % ILI above 2.3, after week 30	Calculation from ILINet (31)	42	24.4	97.4	10.3		
pw4	Percentage of weeks with number of cases above 1/3 of the peak	Calculation from ILINet (31)	22.4	14.2	69.2	10.3		
pw5	Number of weeks from peak until ILI dropped below 1/3 of the peak	Calculation from ILINet (31)	5.9	3.9	21	2		
pw6	Number of weeks from peak until ILI dropped below 1/2 of the peak	Calculation from ILINet (31)	4	2.5	16	2		
Group 13: H1N1 CDC program								
alf	Allocated percent of population by 2/10/2010	Report CDC (143)	0.48	0.06	0.71	0.41		
aln	Allocated percent of population by 11/12/2009	Report CDC (143)	0.14	0.01	0.15	0.13		
dp1	Driving mileage from state capital to assigned depot	Calculation from report CDC(3)	742.4	612.3	3427	10		
dp2	Depot is California (1) or not (0)	Report CDC (3)	0.2	0.4	1	0	10	41
Group 14: H1N1 State program design								
um1	Allocation decision level: 1- State, 2- Mix, 3- Local	University of Michigan (32)	1.75	0.93	3	1		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
um2	Vaccine in public VS private sector. 1 = More Public, 2 = More Private, 3 = Fairly Even mix, 4 = Unsure	University of Michigan (32)	2.43	1.26	4	1		
um4	Estimated percentage of VFC providers registered and participating	University of Michigan (32)	75.79	18.89	100	24		
um8	Who approved private H1N1 providers? 0- State, 1- Locals	University of Michigan (32)	0.34	0.47	1	0	17	34
um9	Estimated % of VCF providers who gave H1N1 vaccine: 0- DK or less than 50%, 1- 50% to 75%, 2- more than 75%	University of Michigan (32)	1.29	0.8	2	0		
um10	Who determined allocation amount to private providers? 0- State, 1- Local	University of Michigan (32)	0.4	0.49	1	0	20	30
um111	Locals make transfer/redistribution to small providers	University of Michigan (32)	0.68	0.47	1	0	34	16
um112	State makes transfer/redistribution to small providers	University of Michigan (32)	0.26	0.44	1	0	13	37
um113	Third parties make transfer/redistribution to small providers	University of Michigan (32)	0.16	0.37	1	0	8	42
um12	Could state track vaccine to final site? 1- Yes, 0- No or U	University of Michigan (32)	0.4	0.49	1	0	20	30
um131	Early allocation strategy (first few weeks) was some per county? 1- Yes, 0- No	University of Michigan (32)	0.16	0.37	1	0	8	42
um132	Early allocation strategy (first few weeks) was pro rata? 1- Yes, 0- No	University of Michigan (32)	0.56	0.5	1	0	28	22
um133	Early allocation strategy (first few weeks) was focused on target populations? 1- Yes, 0- No	University of Michigan (32)	0.26	0.44	1	0	13	37
um14	School-based clinic strategy: 1- School Focus, 0- No	University of Michigan (32)	0.3	0.46	1	0	15	35
umL	1- if locals have the allocation decision, 0- if not	University of Michigan (32)	0.33	0.47	1	0	17	34
umm	1- if state and locals shared allocation decisions, 0- if not	University of Michigan (32)	0.08	0.27	1	0	4	47

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
ums	1- if state has the allocation decision, 0- if not	University of Michigan (32)	0.59	0.49	1	0	30	21
Group 15: H1N1 State program execution								
ag	Cumulative of provider agreements to 12/16/09, per thousand population	Report CDC (33)	0.37	0.1	0.52	0.18		
o1	Average number of days between allocation and order	Calculation from CDC report (2)	4.7	2.2	9.2	0.5		
o2	Average number of daily orders per hundred thousand population	Calculation from CDC report (2)	1.6	1.5	6	0.1		
pd	Total Public Doses Oct-Feb divided by Estimated People Vaccinated	Calculation from CDC report (35)	39.6	20.3	98.9	11.9		
ph	H1N1 Vaccine Doses Distributed or Administered to Date from Large Pharmacy Chains / Retail-Based Clinics to States as of January 29, 2010	CDC Report (57)	10	6.6	30.1	0		
um3	Status of School Clinics by Oct. 27: 1- Some clinics held, 2- Holding doses for school clinics, 3- Postponed until greater vaccine availability	University of Michigan (32)	2.43	0.81	3	1		
um5	Coverage expanded to general population by Dec. 4, 2009: 1- Yes, 0- No	University of Michigan (32)	0.42	0.49	1	0	20	28
um6	Coverage expanded to general population by Dec. 18, 2009: 1- Yes, 0- No	University of Michigan (32)	0.82	0.38	1	0	42	9
um7	Retail vaccination is used by Dec. 18, 2009: 2- Widely, 1- Moderated or in certain counties, 0- Not used	University of Michigan (32)	1.56	0.73	2	0		
Group 16: H1N1 Shipments								
o3	Average number of Shipments per day from October to Dec. 9, 2009	Calculation from CDC report (2)	40.2	45.3	223.6	0.5		

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
o4	Average number of doses per shipment from October to Dec. 9, 2009	Calculation from CDC report (2)	773	1033.4	6365.7	222.4		
o5	1/o3	Calculation from CDC report (2)	0.13	0.34	1.91	0.00 ¹		
o6	Percentage of days with shipments from October to Dec. 9, 2009	Calculation from CDC report (2)	53.7	12.1	71.6	19.4		
o7	Average days from allocation to shipment of vaccine	Calculation from CDC report (2)	6.3	2.3	12.5	2.1		
sh1	Number of Shipments per thousand population from October to Dec. 9, 2009	CDC Shipments Report (Calculation) (2)	0.5	0.25	1.18	0.02		
sh2	Maximum number of ship-to sites per state per thousand population	Report CDC (Calculation) (3)	0.5	0.1	0.74	0.00 ²		
sh3	Number of ship-to sites by unique combination of name and address PTP	Report CDC (Calculation) (2)	0.16	0.08	0.3	0.0 ³		
sh1_3	sh1/sh3	CDC Shipments Report (Calculation) (2)	3.98	2.96	19.64	1.56		
sh4	Average number of days between orders and their shipment	Report CDC (Calculation) (2)	3.02	1.06	8.14	1.44		
sh7	Average of the absolute ratio of the differences between the county per capita doses shipped and the average per capita doses shipped to the state, by the state per capita doses shipped average.	Report CDC (Calculation) (2)	0.58	0.87	6.29	0		

¹ Rounded value

² Rounded value

³ Rounded value

Variable	Description	Reference	Average	Std. Dev.	Max	Min	Number of "yes" (binary)	Number of "no" (binary)
sh8	Percentage of doses sent to primary care, MDs, counties, hospitals, urgent care, clinics, or pharmacies.	Report CDC (Calculation) (2)	66.36	20.05	99.37	4.69		
sh9	Percentage of doses sent to universities, military, jails, government, companies and VAs	Report CDC (Calculation) (2)	2.15	2.43	13.07	0.1		
sh10	Percentage of doses sent to long care, internists, specialists, nursing homes, and children	CDC Shipments Report (Calculation) (2)	7.8	6	27.44	0		
sh11	Percentage of doses sent to internists or specialists	CDC Shipments Report (Calculation) (2)	1.17	1.08	6.07	0		
sh12	Percentage of doses sent to counties	CDC Shipments Report (Calculation) (2)	47.76	25.83	99.37	0		
sh13	Percentage of doses sent to long care and VAs	CDC Shipments Report (Calculation) (2)	0.75	0.52	2.86	0.06		
sh14	Percentage of doses sent to children	CDC Shipments Report (Calculation) (2)	6.69	5.25	22.16	0		
sh15	Percentage of doses sent to primary care	CDC Shipments Report (Calculation) (2)	2.93	2.3	10.44	0		
sh16	Percentage of doses sent to counties, children and primary care	CDC Shipments Report (Calculation) (2)	57.34	21.93	99.37	13.15		
sh17	Number of Ship to sites (by Name and Address) in November with respect to the number in October	CDC Shipments Report (Calculation) (2)	0.97	0.88	4.36	0		
sh18	Shipments per ship to site sent to other than children or county	CDC Shipments Report (Calculation) (2)	2.63	2.05	15	0.11		

APPENDIX B

VARIABLES HIGHLY CORRELATED WITH FACTORS IN THE ADULT MODEL

The table below provides examples of other variables highly correlated with those factors in the final adult's model.

Variable	Description	Examples of highly correlated variables ('+' is > 0.5, '-' is < -0.5)
Indep1	Percent of Women Age 18 and Older Who Report Having Had a Pap Smear Within the Last Three Years, 2008	+ Percent of women who had a mammogram within the last two years + Percent of Women who had a colorectal cancer screening + Child had one or more preventive medical care visits - Percentage of population < 18 years old - Percentage of doses sent to counties
Indep2	Resident population: Hispanic or Latino Origin, percent (July 1 2009 - estimate)	+ Total state population + Chronic liver disease mortality - Proportion of children with medical home - Percentage of white population - Annual retail prescription drugs per capita - Child has health insurance
Indep3	Average days from allocation to shipment of vaccine	+ Average days between allocation and orders + Average number of days between order and shipment - Shipments per ship-to site
Indep4	Percentage of weeks with % ILI above 2.3, after week 30	+ Mortality with end-stage renal disease + Percentage reporting not seeing a doctor in the past 12 months because of cost + Percentage of black population - High educational attainment - Proportion of children with higher insurance - Proportion of children with medical home - Good health status
Indep5	Seasonal influenza Coverage for non-high risk adults 18 -- 49 yrs. on the 2007-2008 season	+ Other previous seasonal influenza coverage rates
Indep6	Maximum number of ship-to sites per state per thousand population	+ Resident population 65 years old and over + Medicare enrollment as percent of total population - State's area
Indep7	Percentage of doses categorized as sent to internists or specialists	+ Percentage of doses sent to long care, internists, specialists, nursing homes, and children + Number of ship-to-sites + Percentage of doses sent to children - Percentage of doses sent to counties

APPENDIX C

VARIABLES HIGHLY CORRELATED WITH FACTORS IN THE CHILDREN AND HIGH-RISK ADULT MODELS

The table below provides examples of other variables highly correlated with those factors in the final adult's model.

Variable	Description	Examples of highly correlated variables ('+' is > 0.5, '-' is < -0.5)
Indep1	Percent of Women Age 18 and Older Who Report Having Had a Pap Smear Within the Last Three Years, 2008	+ Percent of women who had a mammogram within the last two years + Percent of Women who had a colorectal cancer screening + Child had one or more preventive medical care visits - Percentage of population < 18 years old - Percentage of doses sent to counties
Indep6	Maximum number of ship-to-sites per state per thousand population	+ Resident population 65 years old and over + Medicare enrollment as percent of total population - Area of the state
Indep8	Percentage Reporting Not Seeing a Doctor in the Past 12 Months Because of Cost	+ Percent people in poverty + General health condition fair/poor + Deaths by stroke per 100,000 population - Have any kind of healthcare coverage - Percent of adults with dental visit in a year - Educational attainment at least high school
Indep9	Underserved Population Living in Primary Care Health Professional Shortage Areas, as of September, 2008	+ Percent people in poverty + Chronic liver disease mortality - Percent children with private insurance - Peak week for number of ILI cases after week 30
Indep10	Resident population under 18 years, percent (July 1 - estimate) 2008	+ Percent of births to population + Average persons per household + Square root of state's area - Per capita healthcare expenditure by state - Number of healthcare practitioners - Percent of women 40+ with mammogram
Indep11	Resident population: American Indian and Alaska Native alone, percent (July 1 - estimate) 2008	+ Average days between shipments + Area of the state + Average number of doses per shipment
Indep12	Total Public Doses Oct-Feb divided by Estimated People Vaccinated	- Percentage of days with shipments
Indep13	H1N1 Vaccine Doses Distributed or Administered to Date from Large Pharmacy Chains / Retail-Based Clinics to States as of January 29, 2010	(no high correlations were found)
Indep14	Seasonal Flu Coverage for adults 18 -- 49 yrs on the 2007-2008 season	+ Previous seasons vaccination coverage
Indep15	Natural logarithm of the ratio of number of shipments to number of ship-to-sites	+ Percentage of doses sent to counties + Average days between shipments - Number of ship-to sites actually used

Variable	Description	Examples of highly correlated variables ('+' is > 0.5, '-' is < -0.5)
Indep16	Percentage of doses sent to primary care, MDs, counties, hospitals, urgent care, clinics, or pharmacies.	<ul style="list-style-type: none"> + Square root of state's area - Percent of women 40+ with mammogram - Adults 50+ with sigmoidoscopy or colonoscopy - Number of ship-to sites actually used - Percent of adults with dental visit in a year - Average travel time to work - Educational attainment at least high school
Indep17	Total number of cars, trucks and buses per capita	<ul style="list-style-type: none"> - Average travel time to work - Percentage of children living with their grandparents
Indep18	Coverage expanded to general population by Dec. 4, 2009: 1- Yes, 0- No	+ Deaths by stroke per 100,000 population
Indep19	School-based clinic strategy: 1- School Focus, 2- No	<ul style="list-style-type: none"> + Coverage expanded to general population by Dec. 4, 2009 + State makes transfer/redistribution to small providers - Determination of allocation amount to private providers made locally vs. state
Indep20	Third parties make transfer/redistribution to small providers	(no high correlations were found))

APPENDIX D

COMPARISON FOR OBESITY ESTIMATION USING LINEAR

REGRESSION

As a mean to validate the use of logistic regression versus linear regression, we repeated the methodology presented in Chapter 4 using linear regression modeling, for the adult obesity prevalence at a county level in Georgia. We started with the same set of variables used for the adult validation (Appendix F), and performed the variable selection process to build a significant model. The resulting linear regression model is presented in the Table below.

Table D.1: Linear regression model for adults

Age group		Intercept	BNH	HIS	ONH	EDU	HHS	AGE	Gender	ERA
18-24	Estimate	25.71	1.65	0	-1.38	-0.61	0	0.49	0.58	0
	SE	0.45	0.29	0	0.5	0.12	0	0.05	0.23	0
	p-value	0.0000	0.0000	NA	0.0076	0.0000	NA	0.0000	0.0136	NA
25-34	Estimate	28.27	2.82	1.18	-1.02	-0.41	0	0.1	0	0
	SE	0.51	0.34	0.33	0.42	0.1	0	0.04	0	0
	p-value	0.0000	0.0000	0.0007	0.0183	0.0002	NA	0.0106	NA	NA
35-44	Estimate	29.41	2.07	0	-2	-0.42	0	0	0	0.24
	SE	0.43	0.29	0	0.46	0.09	0	0	0	0.08
	p-value	0.0000	0.0000	NA	0.0000	0.0000	NA	NA	NA	0.0023
45-54	Estimate	29.91	1.97	0	-1.86	-0.28	0	0	0	0
	SE	0.42	0.29	0	0.48	0.11	0	0	0	0
	p-value	0.0000	0.0000	NA	0.0002	0.0125	NA	NA	NA	NA
55-64	Estimate	29.88	1.45	0	-3.02	-0.37	0	0	0.64	0
	SE	0.56	0.34	0	0.61	0.11	0	0	0.25	0
	p-value	0.0000	0.0000	NA	0.0000	0.0014	NA	NA	0.0114	NA
>=65	Estimate	29.6	1.4	0	-3.34	-0.22	0	-0.15	0	0.26
	SE	0.38	0.26	0	0.61	0.08	0	0.02	0	0.07
	p-value	0.0000	0.0000	NA	0.0000	0.0044	NA	0.0000	NA	0.0007

Using simulation, as in Chapter 4, we obtain the adult obesity estimation using linear regression to predict BMI of the individual. To decide whether an individual is obese or not, we compare the simulated BMI with 30 (BMI limit used to determine obesity in adults), and classifying the individual as obese if the estimation is greater than this limit value. A graph comparing the confidence interval to the estimation conducted by the National Diabetes Surveillance Service of the CDC in 2007 and 2009 is shown below.

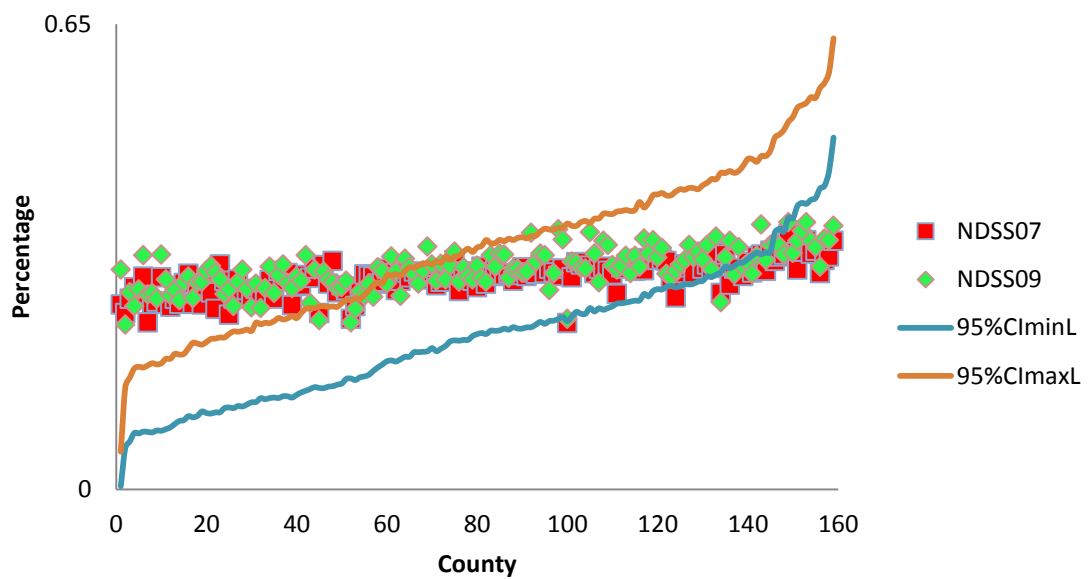


Figure D.1: 95% Confidence interval for obesity adult estimation at a county level in Georgia, using linear models, and compared to estimations from CDC.

When compared to the county level values obtained by the CDC surveys, we observe that the linear model overestimates high obesity levels and underestimates low obesity prevalence. While the average linear estimation of prevalence ranges between 3%

and 56%, the surveyed values range between 23% and 37%. The county-level adult estimations of the logistic regression approach presented in chapter 4, and adjusted by the state average, are compared to the CDC's county level estimates in the figure below.

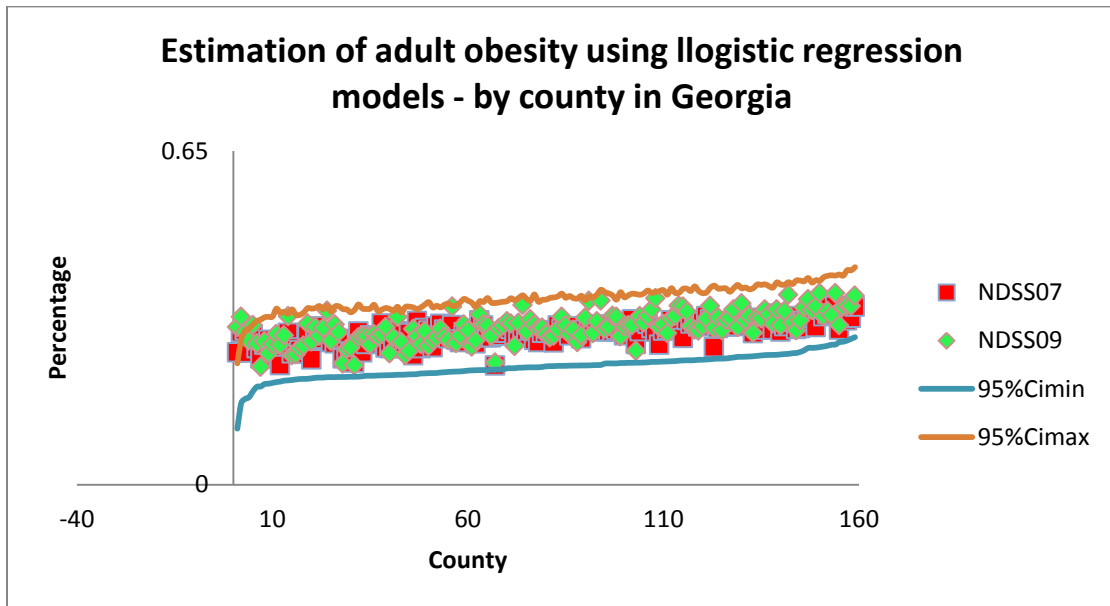


Figure D.2: 95% Confidence interval for obesity adult estimation at a county level in Georgia, using logistic models, adjusting by state average, and compared to estimations from CDC.

We observe that most of the CDC's county level estimations fall within the 95% confidence interval of the logistic regression estimation adjusted by the state average. We conclude that the use of logistic regression in our simulated estimation offers a better approach than linear regression.

APPENDIX E

LIST OF CENSUS TABLES USED TO OBTAIN INFORMATION OF SMALL GEOGRAPHIC AREAS

A. For Household Size:

From Census 2010 summary file 1, we created files with Household size information for each of the following races/ethnicities:

- a. Total population
- b. Hispanic
- c. White Non-Hispanic
- d. Black

Each of these files contained the following areas, in the given order:

- a. US
- b. State (GA)
- c. Counties in the state
- d. Census tracts in the state.

Tables used are:

PCT0280001	Households: Total
PCT0280002	Households: Family households
PCT0280003	Households: Family households; 2-person household
PCT0280004	Households: Family households; 3-person household
PCT0280005	Households: Family households; 4-person household
PCT0280006	Households: Family households; 5-person household
PCT0280007	Households: Family households; 6-person household
PCT0280008	Households: Family households; 7-or-more-person household

B. For Education Level:

From the American Community Survey, we created files with projections for 2010 Educational attainment information for each of the following races/ethnicities:

- a. Total population
- b. Non-Hispanic
- c. Hispanic
- d. White Non-Hispanic
- e. Black Non-Hispanic.

Each of these files contained the following areas, in the given order:

- f. US
- g. State (GA)
- h. Counties in the state
- i. Census tracts in the state

The table used is B15001.

C. For Age:

1. From Census 2010 summary file 1, we created excel files with Age information for each of the following races/ethnicities:

- a. Total population
- b. Non-Hispanic
- c. Hispanic
- d. White Non-Hispanic
- e. Black Non-Hispanic.

Each of these files contained the following areas, in the given order:

- f. US
- g. State (GA)
- h. Counties in the state
- i. Census tracts in the state

The used tables are:

PCT0120005	Total population: Male; 2 years
PCT0120006	Total population: Male; 3 years
PCT0120007	Total population: Male; 4 years
PCT0120008	Total population: Male; 5 years
PCT0120009	Total population: Male; 6 years
PCT0120010	Total population: Male; 7 years
PCT0120011	Total population: Male; 8 years
PCT0120012	Total population: Male; 9 years
PCT0120013	Total population: Male; 10 years
PCT0120014	Total population: Male; 11 years
PCT0120015	Total population: Male; 12 years
PCT0120016	Total population: Male; 13 years
PCT0120017	Total population: Male; 14 years
PCT0120018	Total population: Male; 15 years
PCT0120019	Total population: Male; 16 years
PCT0120020	Total population: Male; 17 years
PCT00120109	Total population: Female; 2 years
PCT00120110	Total population: Female; 3 years
PCT00120111	Total population: Female; 4 years
PCT00120112	Total population: Female; 5 years
PCT00120113	Total population: Female; 6 years
PCT00120114	Total population: Female; 7 years
PCT00120115	Total population: Female; 8 years
PCT00120116	Total population: Female; 9 years
PCT00120117	Total population: Female; 10 years
PCT00120118	Total population: Female; 11 years
PCT00120119	Total population: Female; 12 years
PCT00120120	Total population: Female; 13 years
PCT00120121	Total population: Female; 14 years
PCT00120122	Total population: Female; 15 years
PCT00120123	Total population: Female; 16 years
PCT00120124	Total population: Female; 17 years

APPENDIX F

MODEL FOR VALIDATION IN ADULT POPULATION

Ages 18 years and above, obtained using NHANES data for years 2001 – 2010.

		Coefficient	b ₀	b ₁	b ₂	b ₃	b ₄	b ₅	b ₆	b ₇	b ₈
		Short Description	<i>Intercept</i>	Black non-Hispanic	Hispanic	Other non-Hispanic	Educa-tion	Household size	Age	Gender	Era (group of NHANES data by years of survey)
		Var. Values		{0,1}	{0,1}	{0,1}	{1,...,5}	{1,...,7}	(x - l) Age in years minus lower limit of age group	{1= male, 2= female}	{1,...,5} 1=(2001-2002) 5=(2009-2010)
Age group in years [l-h]	18-24	Mean	-1.61	0.54	0	0	-0.19	0	0.14	0.33	0
		Std. Dev.	0.2	0.09	0	0	0.05	0	0.02	0.09	0
		p-value	< 0.001	< 0.001	-	-	< 0.001	-	< 0.001	< 0.001	-
	25-34	Mean	-1.4	0.78	0.3	0	-0.11	0	0.04	0.25	0.07
		Std. Dev.	0.21	0.1	0.13	0	0.03	0	0.01	0.08	0.03
		p-value	< 0.001	< 0.001	0.02	-	< 0.001	-	0.006	0.003	0.02
	35-44	Mean	-0.43	0.44	0	-0.67	-0.12	0	0	0	0.07
		Std. Dev.	0.13	0.07	0	0.2	0.03	0	0	0	0.03
		p-value	0.002	< 0.001	-	0.001	< 0.001	-	-	-	0.005
	45-54	Mean	-0.36	0.42	0	-0.67	-0.11	0.05	0	0	0
		Std. Dev.	0.15	0.07	0	0.19	0.03	0.03	0	0	0
		p-value	0.017	< 0.001		< 0.001	0.001	0.048	-	-	-
	55-64	Mean	-0.36	0.35	0	-0.95	-0.09	0	0	0.19	0
		Std. Dev.	0.18	0.1	0	0.27	0.04	0	0	0.08	0
		p-value	0.047	< 0.001	-	< 0.001	0.019	-	-	0.012	-
	≥65	Mean	-0.5	0.51	0	-1.08	-0.08	0	-0.05	0	0.13
		Std. Dev.	0.15	0.09	0	0.32	0.03	0	0.01	0	0.03
		p-value	0.001	< 0.001	-	0.001	0.006	-	< 0.001	-	< 0.001

REFERENCES

1. Centers for Disease Control and Prevention. H1N1 Flu - 2009 H1N1 Vaccination Recommendations. In.
2. Centers for Disease Control and Prevention. FOIA Request, H1N1 Shipments through December 9, 2009. ; 2009 Provided Dec 17, 2009.
3. Centers for Disease Control and Prevention. Vaccine Depos assigned to States.; 2010 (Provided June 16, 2010).
4. Centers for Disease Control and Prevention. H1N1 Flu - Presentation: "2009 H1N1: Overview of a Pandemic, April 2009 - August 2010". In.
5. Centers for Disease Control and Prevention. Seasonal Influenza (Flu) - Final estimates for 2009–10 Seasonal Influenza and Influenza A (H1N1) 2009 Monovalent Vaccination Coverage – United States, August 2009 through May, 2010. In.
6. Centers for Disease Control and Prevention. H1N1 Flu - Allocation and Distribution Q&A. In: Prevention CfDCA, editor.: Centers for Disease Control and Prevention; 2009.
7. Centers for Disease Control and Prevention. Interim results: state-specific influenza A (H1N1) 2009 monovalent vaccination coverage - United States, October 2009-January 2010. MMWR Morbidity and Mortality Weekly Report 2010;59(12):363-8.
8. Chapman GB, Coups EJ. Predictors of influenza vaccine acceptance among healthy adults. Preventive Medicine 1999;29(4):249-262.
9. Phillips-Caesar E, Coady MH, Blaney S, Ompad DC, Sisco S, Glidden K, et al. Predictors of influenza vaccination in an urban community during a national shortage. Journal of Health Care for the Poor and Underserved 2008;19(2):611-624.
10. Shen Z-JM, Daskin MS. Trade-offs between customer service and cost in integrated supply chain design. Manufacturing & Service Operations Management 2005;7(3):188-207.
11. Frew PM, Hixson B, del Rio C, Esteves-Jaramillo A, Omer SB. Acceptance of Pandemic 2009 Influenza A (H1N1) Vaccine in a Minority Population: Determinants and Potential Points of Intervention. Pediatrics 2011;127:S113-S119.

12. Frew PM, Painter JE, Hixson B, Kulb C, Moore K, Del Rio C, et al. Factors mediating seasonal and influenza A (H1N1) vaccine acceptance among ethnically diverse populations in the urban south. *Vaccine* 2012.
13. Galarce EM, Minsky S, Viswanath K. Socioeconomic status, demographics, beliefs and A(H1N1) vaccine uptake in the United States. *Vaccine* 2011;29(32):5284-5289.
14. Painter JE, Gargano LM, Sales JM, Morfaw C, Jones LM, Murray D, et al. Correlates of 2009 H1N1 influenza vaccine acceptability among parents and their adolescent children. *Health Educ Res* 2011;26(5):751-60.
15. Jones C, Taylor G, Whittle J, Evans D, Trotter D. Water fluoridation, tooth decay in 5 year olds, and social deprivation measured by the Jarman score: analysis of data from British dental surveys. *BMJ* 1997;315(7107):514-517.
16. Jones C, Worthington H. Water fluoridation, poverty and tooth decay in 12-year-old children. *Journal of Dentistry* 2000;28(6):389.
17. U. S. Census Bureau. National and States Population Estimates. In: Bureau USC, editor.: U. S. Census Bureau; 2009.
18. U. S. Census Bureau. Population Estimates - Quick Tables. In: U. S. Census Bureau; 2008.
19. Volscho T. Measures of Income Distribution in the United States, 1970-2000: Department of Sociology, Anthropology, and Social Work, CUNY – College of Staten Island; 2009.
20. James CV, Salganico A, Thomas M, Ranji U, Lillie-Blanton M, Wyn R. Putting women's health care disparities on the map: The Henry J. Kaiser Family Foundation; 2009.
21. National Automobile Dealers Association. Report on Average age and Number of Vehicles. In: National Automobile Dealers Association; 2006.
22. National Association of County & City Health Officials. 2008 National Profile of Local Health Departments. Washington D. C.; 2009.
23. U. S. Department of Labor - Bureau of Labor Statistics. Create customized tables. In: U. S. Department of Labor - Bureau of Labor Statistics; 2010.
24. American Medical Association. Physician Characteristics and distribution in the US. Chicago IL;; 2006.

25. The Henry J. Kaiser Family Foundation - Statehealthfacts.org. 50 State Comparisons. In.
26. Centers for Disease Control and Prevention. Behavioral Risk Factor Surveillance System - Prevalence and Trends Data. In; 2008.
27. Centers for Disease Control and Prevention. Seasonal Influenza (Flu). Influenza Vaccination Coverage. In.
28. United Health Foundation. America's Health Rankings. In.
29. Centers for Disease Control and Prevention. Instructions for Preparing and Submitting Budget Period 10 Funding Applications. In; 2009.
30. Centers for Disease Control and Prevention. Public Health Emergency Response (PHER) Phase III. In; 2010.
31. Centers for Disease Control and Prevention. Influenza-like Illness Surveillance Program (ILINet); 2010 (Provided July 7, 2010).
32. Clark S. H1N1 Vaccine Implementation Simultaneous Tracking Project - Summary of Findings weeks 09-28-09 to 10-2-09, 10-05-09 to 10-09-09, 10-19-09 to 10-23-09, 10-26-09 to 10-30-09, 11-02-09 to 11-06-09, 11-09-09 to 11-13-09, 11-23-09 to 11-28-09, 11-30-09 to 12-04-09, 12-14-09 to 12-18-09, 01-04-10 to 01-08-10, and 01-11-10 to 01-15-10: University of Michigan; 2010.
33. Centers for Disease Control and Prevention. Cumulative of provider agreements to December 16 2009 per thousand population. ; 2010 Provided May 3.
34. Koonin LM, Beauvais DR, Shimabukuro T, Wortley PM, Palmier JB, Stanley TR, et al. CDC's 2009 H1N1 Vaccine Pharmacy Initiative in the United States: Implications for Future Public Health and Pharmacy Collaborations for Emergency Response Disaster Medicine and Public Health Preparedness 2011;5(4):253 - 255.
35. Centers for Disease Control and Prevention. State data on public doses of H1N1 vaccine. ; 2010 Provided May 6, 2010.
36. George EI. The variable selection problem. Journal of the American Statistical Association 2000;95(452):1304-1308.
37. Rawlings JO, Pantula SG, Dickey DA. Applied regression analysis: a research tool. 2nd ed. New York, NY: Springer; 1998.
38. Park S. Penalized method based on representatives and nonparametric analysis of gap data. Atlanta, GA: Georgia Institute of Technology; 2010.

39. McGiffen J, Milton E, Carmer SG, Ruesink WG. Diagnosis and treatment of collinearity problems and variable selection in least-squares models. *Journal of economic entomology* 1988;81(5):1265-1270.
40. Cohen B, Ferng YH, Wong-McLoughlin J, Jia HM, Morse SS, Larson EL. Predictors of flu vaccination among urban Hispanic children and adults. *Journal of Epidemiology and Community Health* 2012;66(3):204-209.
41. Linn ST, Guralnik JM, Patel KV. Disparities in Influenza Vaccine Coverage in the United States, 2008. *Journal of the American Geriatrics Society* 2010;58(7):1333-1340.
42. Ortiz JR, Neuzil KM, Victor JC, Aitken ML, Goss CH. Predictors of Influenza Vaccination in the Cystic Fibrosis Foundation Patient Registry, 2006 Through 2007. *Chest* 2010;138(6):1448-1455.
43. Bish A, Yardley L, Nicoll A, Michie S. Factors associated with uptake of vaccination against pandemic influenza: A systematic review. *Vaccine* 2011;29(38):6472-6484.
44. Centers for Disease Control and Prevention. 2009 H1N1 and Seasonal Influenza and Hispanic Communities: Questions and Answers. In; 2010.
45. Centers for Disease Control and Prevention. Interim Results: Influenza A (H1N1) 2009 Monovalent Vaccination Coverage --- United States, October--December 2009. *MMWR Morbidity and Mortality Weekly Report* 2010;59(02):44-48.
46. Centers for Disease Control and Prevention. Use of Influenza A (H1N1) 2009 Monovalent Vaccine: Recommendations of the Advisory Committee on Immunization Practices (ACIP). *MMWR* 2009;58(RR10):1-8.
47. Davila-Payan C, Swann J, Wortley PM. System factors to explain H1N1 state vaccination rates for adults in US emergency response to pandemic. *Vaccine* (0).
48. Lu PJ, Callahan DB, Ding H, Euler GL. Influenza A (H1N1) 2009 Monovalent Vaccination Among Adults with Asthma, US, 2010. *American Journal of Preventive Medicine* 2011;41(6):619-626.
49. Sammon CJ, McGrogan A, Snowball J, de Vries CS. Factors associated with uptake of seasonal and pandemic influenza vaccine among clinical risk groups in the UK: An analysis using the General Practice Research Database. *Vaccine* 2012;30(14):2483-2489.
50. SteelFisher GK, Blendon RJ, Bekheit MM, Mitchell EW, Williams J, Lubell K, et al. Novel pandemic A (H1N1) influenza vaccination among pregnant women: motivators and barriers. *American Journal of Obstetrics and Gynecology* 2011;204(6):S116-S123.

51. Durham MD, Buchacz K, Armon C, Patel P, Wood K, Brooks JT, et al. Rates and correlates of influenza vaccination among HIV-infected adults in the HIV Outpatient Study (HOPS), USA, 1999-2008. *Preventive Medicine* 2011;53(1-2):89-94.
52. Singleton J, Santibanez T, Lu P, Ding H, Euler G, Armstrong G, et al. Interim results: influenza A (H1N1) 2009 monovalent vaccination coverage-United States, October-December 2009. *MMWR* 2010;59:44-8.
53. Volscho T. Measures of Income Distribution in the United States, 1970-2000: Department of Sociology, Anthropology, and Social Work, CUNY – College of Staten Island; 2009.
54. James CV, Salganico A, Thomas M, Ranji U, Lillie-Blanton M, Wyn R. Putting Women's Health Care Disparities on the Map. In: *Examining Racial and Ethnic Disparities at the State Level: The Henry J. Kaiser Family Foundation*; 2009.
55. Centers for Disease Control and Prevention. Seasonal Influenza (Flu). Influenza Vaccination Coverage Levels. In.
56. The Child and Adolescent Health Measurement Initiative. National Survey of Children's Health - Physical and Dental Health. In.
57. Centers for Disease Control and Prevention. H1N1 Vaccine Doses Distributed or Administered to Date from Large Pharmacy Chains / Retail-Based Clinics to States as of January 29, 2010.; 2010 (Provided May 20, 2010).
58. Institute for Statistics and Mathematics. R project for statistical computing. In.
59. Kleinbaum DG. Applied regression analysis and multivariable methods: CengageBrain. com; 2007.
60. Althoff KN, Anastos K, Nelson KE, Celentano DD, Sharp GB, Greenblatt RM, et al. Predictors of reported influenza vaccination in HIV-infected women in the United States, 2006-2007 and 2007-2008 seasons. *Prev Med* 2010;50(5-6):223-9.
61. Clark SJ, Cowan AE, Wortley PM. Obstetricians and the 2009–2010 H1N1 Vaccination Effort: Implications for Future Pandemics. *Maternal and Child Health Journal* 2012:1-6.
62. Centers for Disease Control and Prevention. 2009 H1N1 Influenza Vaccine - Dose Spacing for Children 6 Months through 9 Years of Age. In; 2009.
63. Davila-Payan C, Swann J, Wortley PM. System factors to explain H1N1 state vaccination rates for adults in US emergency response to pandemic. In: *Vaccine*; 2013.

64. Centers for Disease Control and Prevention. Winnable Battles. In; 2012.
65. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of Obesity and Trends in Body Mass Index Among US Children and Adolescents, 1999-2010. *Jama-Journal of the American Medical Association* 2012;307(5):483-490.
66. Centers for Disease Control and Prevention. Growth Charts (BMI-for-age). In; 2009.
67. Centers for Disease Control and Prevention. Overweight and Obesity - Basics About Childhood Obesity. In.
68. Daniels SR. The consequences of childhood overweight and obesity. *Future of Children* 2006;16(1):47-67.
69. McCall A, Raj R. Exercise for Prevention of Obesity and Diabetes in Children and Adolescents. *Clinics in Sports Medicine* 2009;28(3):393-+.
70. Berkey CS, Rockett HRH, Gillman MW, Colditz GA. One-year changes in activity and in inactivity among 10-to 15-year-old boys and girls: Relationship to change in body mass index. *Pediatrics* 2003;111(4):836-843.
71. Rollins BY, Belue RZ, Francis LA. The Beneficial Effect of Family Meals on Obesity Differs by Race, Sex, and Household Education: The National Survey of Children's Health, 2003-2004. *Journal of the American Dietetic Association* 2010;110(9):1335-1339.
72. Campbell K, Hesketh K, Crawford D, Salmon J, Ball K, McCallum Z. The Infant Feeding Activity and Nutrition Trial (INFANT) an early intervention to prevent childhood obesity: Cluster-randomised controlled trial. *Bmc Public Health* 2008;8.
73. Data Resource Center for Child & Adolescent Health. National Survey of Children's Health. In; 2007.
74. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey. In: National Center for Health Statistics, editor. Hyattsville, MD; 2001-2008.
75. Ogden CL, Carroll MD, Kit BK, Flegal KM. Prevalence of Obesity in the United States, 2009-2010. *NCHS data brief* 2012(82):1-8.
76. Centers for Disease Control and Prevention. Youth Risk Behavior Surveillance — United States, 2011. *Morbidity and Mortality Weekly Report (MMWR)* 2012;61(4).
77. Akinbami LJ, Ogden CL. Childhood Overweight Prevalence in the United States: The Impact of Parent-reported Height and Weight. *Obesity* 2009;17(8):1574-1580.

78. Shields M, Gorber SC, Janssen I, Tremblay MS. Obesity estimates for children based on parent-reported versus direct measures. *Health Reports* 2011;22(3).
79. Nihiser AJ, Lee SM, Wechsler H, McKenna M, Odom E, Reinold C, et al. Body Mass Index Measurement in Schools. *Journal of School Health* 2007;77(10):651-671.
80. Centers for Disease Control and Prevention. National Health and Nutrition Examination Survey: Questionnaires, Datasets, and Related Documentation. In; 2010.
81. Gregg EW, Kirtland KA, Cadwell BL, Burrows NR, Barker LE, Thompson TJ, et al. Estimated County-Level Prevalence of Diabetes and Obesity-United States, 2007 (Reprinted from *MMWR*, vol 58, pg 1259-1263, 2009). *Jama-Journal of the American Medical Association* 2010;303(10):933-935.
82. Li WJ, Kelsey JL, Zhang Z, Lemon SC, Mezgebu S, Boddie-Willis C, et al. Small-Area Estimation and Prioritizing Communities for Obesity Control in Massachusetts. *American Journal of Public Health* 2009;99(3):511-519.
83. Malec D, Sedransk J, Moriarity CL, LeClere FB. Small area inference for binary variables in the National Health Interview Survey. *Journal of the American Statistical Association* 1997;92(439):815-826.
84. Malec D, Davis WW, Cao X. Model-based small area estimates of overweight prevalence using sample selection adjustment. *Statistics in Medicine* 1999;18(23):3189-3200.
85. Pfeffermann D. New Important Developments in Small Area Estimation. *Statistical Science* 2013;28(1):40-68.
86. United States Census Bureau. Summary File 1. In; 2010.
87. United States Census Bureau. American Community Survey. In; 2010.
88. Messer SC, Liu X, Hoge CW, Cowan DN, Engel CC. Projecting mental disorder prevalence from national surveys to populations-of-interest - An illustration using ECA data and the U.S. Army. *Social Psychiatry and Psychiatric Epidemiology* 2004;39(6):419-426.
89. Congdon P. A multilevel model for cardiovascular disease prevalence in the US and its application to micro area prevalence estimates. *International Journal of Health Geographics* 2009;8.
90. Zhang Z, Zhang L, Penman A, May W. Using Small-Area Estimation Method to Calculate County-Level Prevalence of Obesity in Mississippi, 2007-2009. *Preventing Chronic Disease* 2011;8(4).

91. Congdon P. Estimating population prevalence of psychiatric conditions by small area with applications to analysing outcome and referral variations. *Health & Place* 2006;12(4):465-478.
92. Judge A, Welton NJ, Sandhu J, Ben-Shlomo Y. Modeling the Need for Hip and Knee Replacement Surgery. Part 2. Incorporating Census Data to Provide Small-Area Predictions for Need With Uncertainty Bounds. *Arthritis & Rheumatism-Arthritis Care & Research* 2009;61(12):1667-1673.
93. Choy M, Switzer P, De Martel C, Parsonnet J. Estimating disease prevalence using census data. *Epidemiology and Infection* 2008;136(9):1253-1260.
94. Heyman RE, Slep AMS. Estimating Prevalences of Sensitive Problems From Nonsensitive Data. *Journal of Interpersonal Violence* 2011;26(2):293-311.
95. Zhang X, Onufrak S, Holt J, Croft J. A Multilevel Approach to Estimating Small Area Childhood Obesity Prevalence at the Census Block-Group Level. *Preventing Chronic Disease* 2013;10(120252).
96. Longjohn M, Sheon AR, Card-Higginson P, Nader PR, Mason M. Learning From State Surveillance Of Childhood Obesity. *Health Affairs* 2010;29(3):463-472.
97. R Core Team. foreign: Read Data Stored by Minitab, S, SAS, SPSS, Stata, Systat, dBase,R package version 0.8-52. In; 2013.
98. Lumley T. survey: analysis of complex survey samples. R package version 3.28-2 In; 2012.
99. Wickham H. ggplot2: elegant graphics for data analysis. In. New York: Springer; 2009.
100. Grow HMG, Cook AJ, Arterburn DE, Saelens BE, Drewnowski A, Lozano P. Child obesity associated with social disadvantage of children's neighborhoods. *Social science & medicine (1982)* 2010;71(3):584.
101. Bethell C, Read D, Goodman E, Johnson J, Besl J, Cooper J, et al. Consistently Inconsistent: A Snapshot of Across- and Within-State Disparities in the Prevalence of Childhood Overweight and Obesity. *Pediatrics* 2009;123:S277-S286.
102. Kock R. The 80/20 Principle-The secret to Achieving More with Less. New York: Doubleday; 1998.
103. United States Census Bureau. Geography Terms and Concepts - Census Tract. In; 2010.

104. Centers for Disease Control and Prevention. Estimated county-level prevalence of diabetes and obesity — United States, 2007. *MMWR Morb Mortal Wkly Rep* 2009;58(45):1259-63.
105. Arkansas Center for Health Improvement, editor. *Year Four Assessment of Childhood and Adolescent Obesity in Arkansas (Fall 2006–Spring 2007)*. Little Rock, AR; 2011.
106. Gaonkar R, Viswanadham N. A conceptual and analytical framework for the management of risk in supply chains. In: *Robotics and Automation, 2004. Proceedings. ICRA'04. 2004 IEEE International Conference on; 2004: IEEE; 2004. p. 2699-2704.*
107. Vljajic JV, van der Vorst JG, Haijema R. A framework for designing robust food supply chains. *International Journal of Production Economics* 2012;137(1):176-189.
108. Dudley JP. Public Health and Epidemiological Considerations For Avian Influenza Risk Mapping and Risk Assessment. *Ecology and Society* 2008;13(2).
109. Boisvert RN, Kay D, Turvey CG. Macroeconomic costs to large scale disruptions of food production: The case of foot-and-mouth disease in the United States. *Economic Modelling* 2012;29(5):1921-1930.
110. Lee HL, Padmanabhan V, Whang SJ. Information distortion in a supply chain: The bullwhip effect. *Management Science* 1997;43(4):546-558.
111. Rong Y, Snyder LV, Shen Z-JM. Bullwhip and reverse bullwhip effects under the rationing game. Available at SSRN 1240173 2008.
112. Wagner SM, Bode C. An empirical examination of supply chain performance along several dimensions of risk. *Journal of Business Logistics* 2008;29(1):307-325.
113. Rong Y, Shen Z-JM, Snyder LV. Pricing during disruptions: a cause of the reverse bullwhip effect. Available at SSRN 1374184 2009.
114. Forrester J. Industrial dynamics: a major breakthrough for decision makers. *Harvard Business Review* 1958;36(4):37-65.
115. Giard V, Sali M. The bullwhip effect in supply chains: a study of contingent and incomplete literature. *International Journal of Production Research* 2013;51(13):3880-3893.
116. Ozelkan EC, Cakanyildirim M. Reverse bullwhip effect in pricing. *European Journal of Operational Research* 2009;192(1):302-312.

117. Zotteri G. An empirical investigation on causes and effects of the Bullwhip-effect: Evidence from the personal care sector. *International Journal of Production Economics* 2013;143(2):489-498.
118. Zhang XL, Burke GJ. Analysis of compound bullwhip effect causes. *European Journal of Operational Research* 2011;210(3):514-526.
119. Bhattacharya R, Bandyopadhyay S. A review of the causes of bullwhip effect in a supply chain. *International Journal of Advanced Manufacturing Technology* 2011;54(9-12):1245-1261.
120. Disney SM, Towill DR. On the bullwhip and inventory variance produced by an ordering policy. *Omega* 2003;31(3):157-167.
121. Csik A, Foldesi P. A bullwhip type of instability induced by time varying target inventory in production chains. *International Journal of Innovative Computing Information and Control* 2012;8(8):5885-5897.
122. Holweg M, Disney S, Holmström J, Småros J. Supply Chain Collaboration:: Making Sense of the Strategy Continuum. *European management journal* 2005;23(2):170-181.
123. Agrawal S, Sengupta RN, Shanker K. Impact of information sharing and lead time on bullwhip effect and on-hand inventory. *European Journal of Operational Research* 2009;192(2):576-593.
124. Cannella S, Ciancimino E. On the Bullwhip Avoidance Phase: supply chain collaboration and order smoothing. *International Journal of Production Research* 2010;48(22):6739-6776.
125. Ouyang YF, Li XP. The bullwhip effect in supply chain networks. *European Journal of Operational Research* 2010;201(3):799-810.
126. de Blasio B, Xue Y, Iversen B, Gran JM. Estimating influenza-related sick leave in Norway: Was work absenteeism higher during the 2009 A (H1N1) pandemic compared to seasonal epidemics? *children* 2012;2:13.
127. Schanzer DL, Zheng H, Gilmore J. Statistical estimates of absenteeism attributable to seasonal and pandemic influenza from the Canadian Labour Force Survey. *Bmc Infectious Diseases* 2011;11.
128. Keogh-Brown MR, Wren-Lewis S, Edmunds WJ, Beutels P, Smith RD. The possible macroeconomic impact on the UK of an influenza pandemic. *Health Economics* 2010;19(11):1345-1360.

129. Sadique MZ, Adams EJ, Edmunds WJ. Estimating the costs of school closure for mitigating an influenza pandemic. *Bmc Public Health* 2008;8.
130. Shi P, Keskinocak P, Swann JL, Lee BY. The impact of mass gatherings and holiday traveling on the course of an influenza pandemic: a computational model. *Bmc Public Health* 2010;10(1):778.
131. Schultz M. Dairy products profile. Agricultural Marketing Resource Center. Available online: http://www.agmrc.org/commodities__products/livestock/dairy/dairy_products_profile.cfm (accessed on 12 April 2011) 2008.
132. ProCon.org. How Milk Gets from the Cow to the Store. In; 2011.
133. Pengyi S, Pinar K, Julie S, Bruce L. The impact of mass gatherings and holiday traveling on the course of an influenza pandemic: a computational model. *Bmc Public Health*;10.
134. U. S. Census Bureau. 1990 Census. In: *American FactFinder*; 1990.
135. U. S. Census Bureau. Summary File 3. In: *American FactFinder*; 2000.
136. StateMaster.com. Total number of licensed drivers. In: *Transportation Statistics > Licensed Drivers > Total number (most recent) by state* 2004.
137. NewsMax.com. Doctors Group Fights Thimerosal Ban In: *NewsMax.com Wires*; 2006.
138. Centers for Disease Control and Prevention. HIV/AIDS Cases of HIV infection and AIDS in the United States and Dependent Areas, 2005. In; Rev. 2007.
139. Centers for Disease Control and Prevention. Chronic Disease Indicators. In: *National Center for Chronic Disease Prevention and Health Promotion*; Rev. 2007.
140. Centers for Disease Control and Prevention. National, State, and Local Area Vaccination Coverage Among Children Aged 19–35 Months — United States, 2008. *MMWR Morbidity and Mortality Weekly Report* 2009;58(33):921-6.
141. Centers for Disease Control and Prevention. National, State, and Local Area Vaccination Coverage Among Adolescents Aged 13--17 Years --- United States, 2008. *MMWR Morbidity and Mortality Weekly Report* 2009;58(36):997-1001.
142. Centers for Disease Control and Prevention. Influenza Vaccination Coverage Among Children Aged 6--23 Months --- United States, 2007--08 Influenza Season. *MMWR Morbidity and Mortality Weekly Report* 2009;58(38):1063-6.

143. Centers for Disease Control and Prevention. Reports on allocated number of vaccines as a percent of total population. ; 2010 Provided June 16.

VITA

CARLO STEFAN DAVILA PAYAN

Carlo was born in Mexico City, Mexico. He received a Bachelor's Degree in Physics Engineering and a Masters in Science in Industrial Engineering, in the track of Production Systems, from ITESM Technology Institute in Mexico, where he also worked as an instructor of physics, mathematics, and industrial engineering subjects. Carlo worked as a manufacturing and delivery planner and then as a manufacturing manager before starting his doctoral studies at Georgia Tech.