THREE ESSAYS OF HEALTHCARE DATA-DRIVEN PREDICTIVE MODELING

by

Zhouyang Lou

A Dissertation

Submitted to the Faculty of Purdue University

In Partial Fulfillment of the Requirements for the degree of

Doctor of Philosophy

School of Industrial Engineering

West Lafayette, Indiana

May 2023

THE PURDUE UNIVERSITY GRADUATE SCHOOL
STATEMENT OF COMMITTEE APPROVAL

Dr. Nan Kong, Co-Chair

Weldon School of Biomedical Engineering

Dr. Zachary Hass, Co-Chair

School of Industrial Engineering

Dr. Hua Cai

School of Industrial Engineering

Dr. Greg Arling

School of Nursing

Dr. Paul Griffin

Department of Industrial and Manufacturing Engineering,

Penn State University

Approved by:

Dr. Young-Jun Sun

Dedicated to my wife, parents, son and my lab mates who always supported my work

ACKNOWLEDGMENTS

I would like to first express my sincere gratitude to my thesis advisors, Dr. Kong and Dr. Hass, for their guidance, encouragement, patience and support throughout my research project. Your expertise and feedback have been invaluable in shaping my work. I am also grateful to each person on my committee. Dr. Arling, thank you for providing me with much help and insights in grant writing and guidance in health services research. Dr. Griffin, thank you for offering me abundant resources and freedom for my research projects. Dr. Cai, thank you for your support and encouragement when I had my first experience teaching a large undergraduate course.

I would like to extend my sincere thanks to my research collaborators throughout my research projects. Special thanks to my friends in my research lab. Carolina, Wayne, Shujing, Xiaoquan, Fangyuan and You have given me feedback, support and encouragement.

Last but not least, this endeavor would not have been possible without unwavering love, understanding and accompany from my family. My wife, Xiaojun, I thank her for her patience and for sacrificing her career to be a full-time housewife to help me and our son Lazarus. She has also given me a lot of spiritual encouragement to reach the finish line. My parents and my parents-in-law, I thank them for being generous, positive, and patient on this long journey.

I also would like to acknowledge that my published essay 1 is reprinted with permission from Elsevier from Clinical Gastroenterology and Hepatology 18.10 (2020): 2305-2314, Cost-effectiveness of different strategies for detecting cirrhosis in patients with nonalcoholic fatty liver disease based on United States health care system, © 2020 by the AGA Institute, and my published essay 2 is reproduced with permission from Springer Nature from Journal of Urban Health 100, 51–62 (2023), The Health and Economic Impact of Using a Sugar Sweetened Beverage Tax to Fund Fruit and Vegetable Subsidies in New York City: A Modeling Study, © 2023 Springer Nature. The research from essay 2 was supported by a grant from the National Heart, Lung, and Blood Institute (R01HL141427) of the National Institutes of Health. The contents of this article are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Predictive modeling in healthcare involves the development of data-driven and computational models which can predict what will happen, be it for a single individual or for an entire system. The adoption of predictive models can guide various stakeholders' decision making in the healthcare sector, and consequently improve individual outcomes and the cost-effectiveness of care. With the rapid development in healthcare of big data and the Internet of Things technologies, research in healthcare decision-making has grown in both importance and complexity. One of the complexities facing those who would build predictive models is heterogeneity of patient populations, clinical practices, and intervention outcomes, as well as from diverse health systems. There are many sub-domains in healthcare for which predictive modeling is useful such as disease risk modeling, clinical intelligence, pharmacovigilance, precision medicine, hospitalization process optimization, digital health, and preventive care. In my dissertation, I focus on predictive modeling for applications that fit into three broad and important domains of healthcare, namely clinical practice, public health, and healthcare system. In this dissertation, I present three papers that present a collection of predictive modeling studies to address the challenge of modeling heterogeneity in health care. The first paper presents a decision-tree model to address clinicians' need to decide among various liver cirrhosis diagnosis strategies. The second paper presents a micro-simulation model to assess the impact on cardiovascular disease (CVD) to help decision makers at government agencies develop cost-effective food policies to prevent cardiovascular diseases, a public-health domain application. The third paper compares a set of data-driven prediction models, the best performing of which is paired together with interpretable machine learning to facilitate the coordination of optimization for hospital-discharged patients choosing skilled nursing facilities. This collection of studies addresses important modeling challenges in specific healthcare domains, and also broadly contribute to research in medical decision-making, public health policy and healthcare systems.

1. INTRODUCTION

Decision-making in healthcare is complex and the range of applications and sub-domains is diverse, such as clinical practice, population health, and healthcare delivery systems. It requires many considerations before arriving at a reasonable course of action in these areas to achieve better patient outcomes.[1] Predictive modeling with mathematical and statistical methods (including machine learning methods) can utilize not only existing healthcare data, but can also be augmented by expert knowledge to improve decision-making processes with the duel goals of improving patient outcomes and reducing healthcare costs. As healthcare data continues to grow and new technologies emerge, predictive modeling becomes increasingly important as the capacity of human cognition to make optimal decisions across multiple criteria when faced with a wealth of input information is taxed to the limit. Predictive models can support human cognition and inform decision-making in a scalable way by managing growing data volumes and information complexity.[2]

Various predictive modeling tools have been devised to assist with the decision-making process, such as identifying patients at high risk for various medical conditions (cardiovascular diseases, liver cirrhosis, diabetes, etc.), predicting the probability of getting some disease in the future, predicting treatment outcomes, and improving the healthcare delivery process.[3] Though predictive models are powerful quantitative tools, they will not replace the qualitative reasoning of healthcare professionals or take over their jobs, but rather estimate objective outcomes to assist their reasoning and decision-making.[4–8]

Predictive models capture and quantify relationships among many factors and processes to predict some future events.[9,10] Model-based and model-free are the two broad types of predictive modeling techniques: An example of a model-based approach is a predictive simulation model based on the domain knowledge of the underlying physical process. Another typical example is multivariate regression methods[11,12] which uses a function (e.g., polynomials, logistic function, etc.) to describe the relationship between the independent and dependent variables. The coefficients of the independent variables capture their impact on the outcomes. Given such a model and a new data instance of the predictor variables, a prediction can be made as the average value of the function

at those predictor variable values. Alternatively, machine-learning techniques and network analytics[13] are examples of model-free data mining tasks. Both types of predictive modeling techniques can be mixed in many complex applications.

Although predictive modeling techniques are becoming increasingly influential in the healthcare sector, many challenges remain. The difficulties are due to the natural complexity of healthcare including the diversity of medical conditions and patient behaviors, the heterogeneity of treatments and outcomes, a diversity of healthcare systems that operate differently, and the subtle differences in study designs, analytical methods, and approaches for collecting, processing, and interpreting healthcare data. Additionally, many key stakeholders may have different interests, priorities, or perspectives when receiving or delivering care from different healthcare domains (i.e., clinical practice, public health, healthcare systems) such as patients, medical practitioners, hospital operators, healthcare insurers, predictive analytics modelers, and government agencies.[10,14,15] For example, from the patient's perspective, the goal is often to be cured of a disease by receiving healthcare services at an affordable cost with personalized recommendations. On the other hand, medical practitioners may be incentivized to deliver care quickly to maximize patients seen or to use treatment from a list of appropriate options with the most considerable impact on their income. Hospital operators must effectively manage and optimize the available healthcare resources to ensure cost-effective hospital operations. Government agencies aim to achieve overall/population-level social welfare with equity at reasonable costs. Thus, no universal framework exists to model, analyze, or compare the performances of various predictive modeling strategies.[16]

Moreover, model transparency and conflicts are an additional layer of challenges to the stakeholders, and they must be adequately addressed before implementing predictive models in the real world. For example, medical practitioners often prefer an intuitive model resembling human perception and decision logic (e.g., a decision rule or tree).[17] Pragmatically, it is essential for clinicians to understand, for example, why some patients are predicted as high risk after being discharged from the hospital while others are not and why specific interventions are being promoted in certain scenarios. Since medical practitioners are held accountable for treating patients using their professional knowledge, it is intolerable for practitioners to trust predictive models which may contain errors blindly.

Finally, many issues related to the data are challenging to predictive modeling in healthcare. First is data access. Due to restrictive government policies for protecting patient health data privacy, a secure data environment must be established, which may include protocols to link data from multiple sources and then de-identify it before analysis. The second is data availability. Patients are not required to report for follow-up or remain in the same part of the healthcare system, meaning many outcomes or metrics are not measured or recorded. Even when they are, many predictor variables, such as health behaviors, may be at best recorded for only a subset of patients. Lastly is data variety. Existing data often comes from multiple sources with diversified formats such as free text fields and images. Many questions about the data (e.g., What data to use? What if we do not have specific data?) must be addressed in developing predictive models for a particular application.[15]

This dissertation aims to cover the challenges above by addressing an overarching question on how to develop appropriate predictive modeling tools for a specific healthcare problem in three important healthcare domains (i.e., clinical practice, public health, and healthcare systems). I use three studies in this dissertation to illustrate sound principles and practices in predictive modeling.

In chapter 2, we develop a predictive model at the clinical practice level to address medical practitioners' needs to assess various liver cirrhosis diagnostic strategies, including existing and proposed new strategies. The predictive model consists of a decision tree and a micro-simulation model to predict patients' health outcomes and associated healthcare costs using each liver cirrhosis diagnostic strategy. The model captures the practical workflow of the Gastroenterology department. This study helps medical practitioners (i.e., gastroenterologists) to compare and identify the most cost-effective diagnostic strategy in daily clinical practice.

In chapter 3, we use cardiovascular disease (CVD) and economic predictive modeling from the public health perspective to assist policymakers at government agencies in evaluating various food policies for preventing cardiovascular diseases cost-effectively. The predictive model uses a disease graph verified by cardiologists to capture individual patient's CVD progression from a heterogeneous population, and logistic regression models to quantify the transition probabilities.

The model is validated against real-world data. Multiple health outcomes and healthcare costs are captured and analyzed to support policymakers' decisions to improve public health.

Chapter 4 focuses on the system consisting of a hospital and multiple skilled nursing facilities (SNF). I developed a readmission prediction tool to predict probabilities of post-acute older-adult patients being re-hospitalized at the time of discharge from a hospital to an SNF. This tool can be used to facilitate the study of the coordination of patient discharge assignment recommendations within the hospital-SNF system.

Lastly, I will discuss my reflection on predictive modeling in healthcare and future research.

2. ESSAY 1: COST EFFECTIVENESS OF DIFFERENT STRATEGIES FOR DETECTING CIRRHOSIS IN PATIENTS WITH NON-ALCOHOLIC FATTY LIVER DISEASE BASED ON UNITED STATES HEALTH CARE SYSTEM

2.1 Introduction

Nonalcoholic fatty liver disease (NAFLD) is a leading cause of chronic liver disease, hepatocellular carcinoma (HCC), and liver transplantation.[18] The extent of hepatic fibrosis is the strongest risk factor for long-term complications; patients with advanced fibrosis are at the highest risk of hepatic decompensation, HCC, and cardiovascular disease.[19] Thus, accurate staging of fibrosis is necessary for proper management of patients with NAFLD, both for estimating prognosis and guiding therapy. While liver biopsy (LB) is the gold standard for staging liver fibrosis and diagnosing cirrhosis, it is invasive, costly, and has technical limitations, including sampling error and variation in histological interpretation,[20] making it impractical as a routine screening or diagnostic tool.

Several non-invasive tests (NITs) are available for staging liver fibrosis, including blood-based biomarkers (Fib-4),[21] and modalities that assess liver stiffness such as vibration controlled transient elastography (VCTE)[22] and magnetic resonance elastography (MRE).[23] Most NITs are good at excluding advanced fibrosis (negative predictive value > 90%), but sensitivity is lower for detecting advanced fibrosis.[24–26] Among the modalities assessing liver stiffness, VCTE is the most widely studied and validated method, yielding good intra- and inter-observer variability and low failure rates.[22] Moreover, its wide availability and easy-to-operate feature make VCTE a potential point-of-care test. Most recently, 2D- and 3D-MRE have emerged as promising noninvasive tests for diagnosing advanced fibrosis in NAFLD patients. MRE and VCTE have high accuracy in identifying advanced fibrosis.[26] The Fib-4 index is a simple and inexpensive blood-based panel that is useful for excluding advanced fibrosis.[21] It has been extensively validated in patients with

NAFLD. Fib-4 cutoffs ≥ 3.25 or ≤ 1.30 have 80% and 90% positive and negative predictive values, respectively, for detecting or excluding advanced fibrosis among patients CLD, including NAFLD.[25,27] Its major limitation is that a high proportion of values fall in the indeterminate range (1.30-3.25).[25,27]

In clinical practice, a sequential combination of two NITs may improve the detection of patients with advanced fibrosis and reduce the number of patients with indeterminate results[28,29]; however, its cost-effectiveness has not been fully explored. Herein, we developed a decision model to quantify the accuracy and costs of various non-invasive strategies for diagnosing cirrhosis in NAFLD patients to understand better which strategies might be considered preferable in clinical practice.

2.2 Methods

2.2.1 Study population

Our study population consists of a hypothetical cohort of middle-aged patients with NAFLD seen in three different settings: (a) in a specialty clinic setting where the prevalence of cirrhosis in NAFLD is estimated to be 2%,[30] (b) in a general population-based setting where the prevalence of cirrhosis is estimated to be 0.27%[31] and tertiary referral centers setting where cirrhosis prevalence is reported to be ≈ 4%.[23,25,26]

2.2.2 Model structures

We constructed a decision model to compare the accuracy and costs associated with liver biopsy and NITs alone and in a sequential combination for the diagnosis of cirrhosis. We considered Fib-4 and liver stiffness measured by VCTE or MRE, and assumed that liver biopsy, as the reference standard, would have the highest sensitivity and specificity for cirrhosis. Test accuracy for each strategy was calculated as follows: (true positive (TP) + true negative (TN))/(total population). Note that we excluded TP and TN numbers from the confirmation test. To evaluate the ensuing clinical events for future costs and outcomes; we attached a microsimulation to each leaf node of the decision tree to emulate subsequent clinical events over a 5-year time horizon. Figures 2.1-3

summarize the decision tree of the nine diagnostic strategies and the transitions in the microsimulation of patients after receiving a diagnosis of cirrhosis.

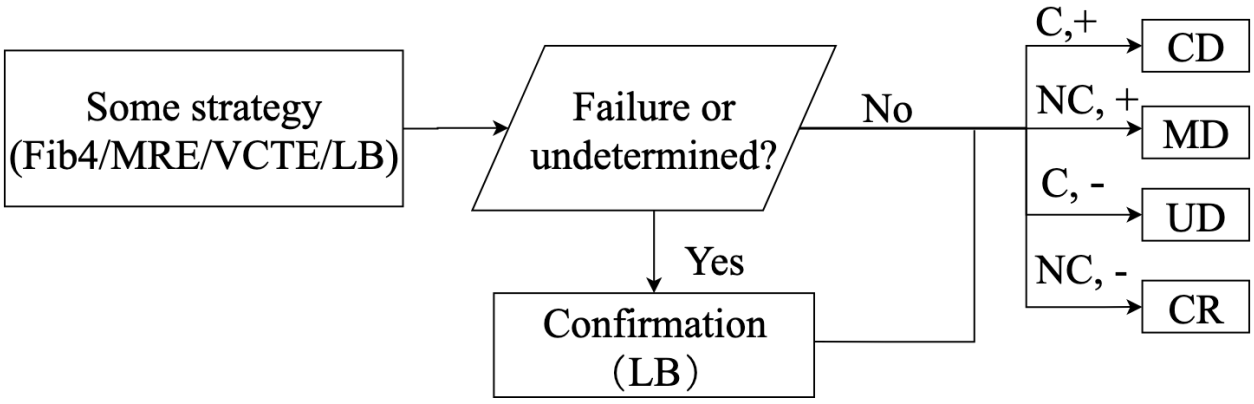


Figure 2.1. Model structure: Decision analytic tree using single tests.

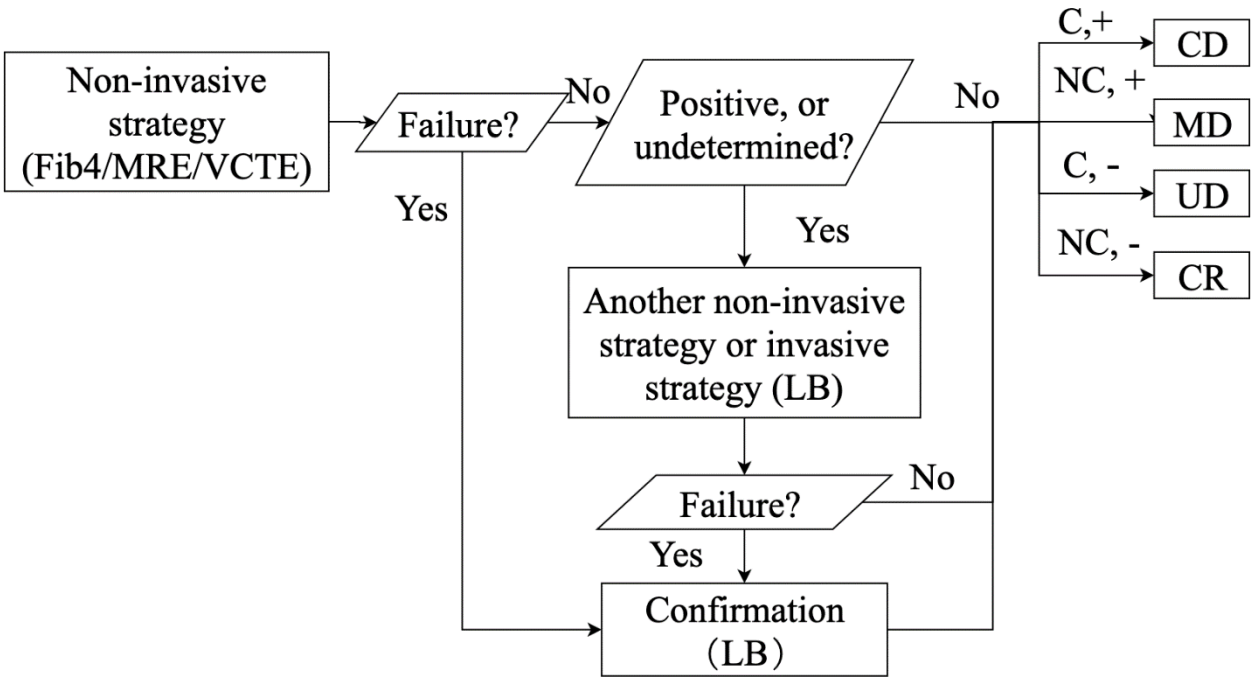


Figure 2.2. Model structure: Decision analytic tree using single tests using sequential tests.

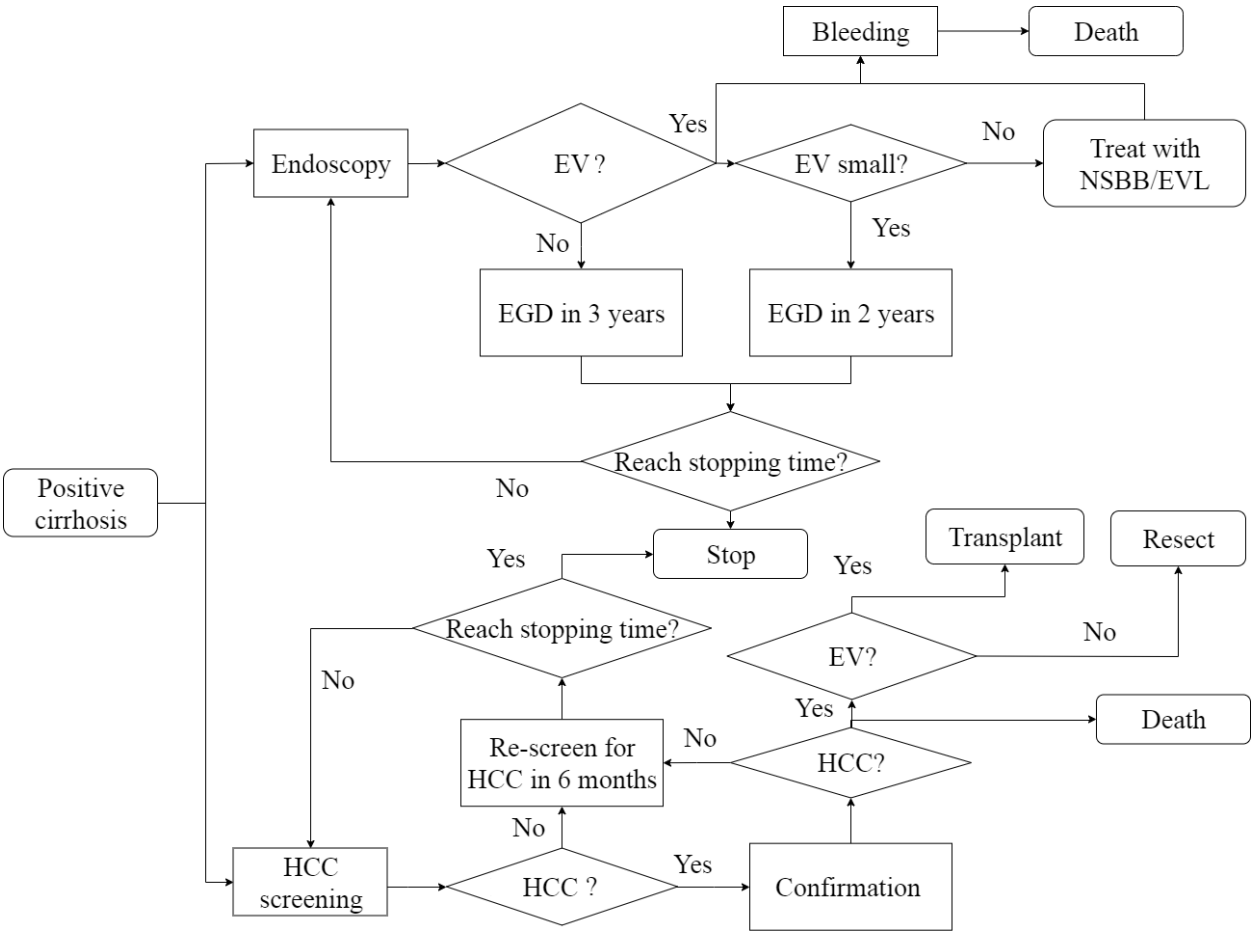


Figure 2.3. Model structure: Microsimulation of patients' assessment after receiving a diagnosis of cirrhosis.

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography; **C**, cirrhosis; **NC**, no cirrhosis; **CD**, correctly diagnosed; **MD**, misdiagnosed; **CR**, correctly ruled out; **UD,** undiagnosed; **HCC**, hepatocellular carcinoma; **EV**, esophageal varices; **EVL**, endoscopic variceal ligation; **OLT**, orthotopic liver transplantation; **NSBB**, nonselective beta-blockers; **LB**, liver biopsy; **EGD**, esophagogastroduodenoscopy.

In the decision-tree portion of the model, we considered each test individually and in clinically appropriate combinations where either stiffness-based method was added sequentially to Fib-4 or where LB was added sequentially to either a stiffness-based method or Fib-4. For all test strategies, a positive test may suggest cirrhosis and a negative test may exclude cirrhosis. Additionally, failure rates for VCTE (7.1%)[22,32] and MRE (4%),[33,34] and indeterminate results for Fib-4 (32%)[22] were included. For combination strategies, a second test was added if the initial test was either positive or indeterminate for Fib-4, or uninterpretable for VCTE/MRE. The result of the second test determined the final diagnosis of cirrhosis. To better reflect the diagnostic practice, we applied the

confirmation test using LB in two cases: 1) indeterminate or failure results occur in Fib-4, MRE, and VCTE only tests, 2) failure results in combination strategies. Note that in calculating the final accuracy, we do not account for TP and TN results from the confirmation test. Finally, a do-nothing approach was included along with all test strategies.

We assumed that all patients with cirrhosis diagnosis underwent periodic guideline-based screening for esophageal varices (EV) and HCC[35,36] over a 5-year period. The disease state and mortality status were updated during the 5-year period. As a result, costs of the periodic screening for EVs and HCC were accumulated over the 5-year period, depending on the simulated disease state and mortality status. For those who tested positive, the costs include: (1) periodic EV screening [every 2 years if EV is small and at 3-year intervals if EVs are absent][36], (2) HCC screening every 6 months using serum α-fetoprotein and ultrasound,[35] (3) primary prophylaxis of EV bleeding with both endoscopic band ligation (EBL) and nonselective beta-blockers (NSBB),[36] (4) outpatient clinic visit and (5) those derived from liver transplantation or HCC-related liver resection. For those who tested negative, the two screening costs as above were excluded.

2.2.3 Model Parameters

We performed a sensitivity analysis by varying the prevalence range from 0.27% to 4%. The prevalence of EV among cirrhotics was estimated to be 34.7%.[37] We considered the annual incidence of EV and HCC to be 2.3% (unpublished data from Indiana University Medical Center) and 2.5%, respectively.[38] The pooled 3-year HCC mortality rate is 49.2% among patients undergoing HCC surveillance vs 72.1% among those without periodic surveillance.[35] We assumed a 16.3% 6-month risk of mortality due to variceal bleeding.[39] We considered a prevalence of 11.5% of large or high-risk varices among compensated NAFLD cirrhotics[40] and assumed that 34% of small or no varices would progress to large varices at 3 years.[41] The pooled 2-year risk of upper gastrointestinal bleeding is 17%[42] and 31%[43] among patients with or without primary prophylaxis including both EBL and/or NSSB. Table 2.1, Figure 2.3 and Appendix A Tables A.1-A.3 summarize some of the above-mentioned parameters used in the model.

Table 2.1. Model parameters, base-case values and ranges.

Parameter	Value	Range	Reference
Prevalence of biopsy-confirmed cirrhosis			
Population-based	0.27%	-	[31]
Community-based	2%		[30]
Prevalence of esophageal varices	34.7%	-	[37]
Prevalence of large or high-risk varices	11.5%	-	[40]
3-year rate of varices progression	34%	-	[41]
Annual incidence of esophageal varices	4.4%	-	-
2-year risk of bleeding with primary prophylaxis	17%	14%-20%	[42]
2-year risk of bleeding without primary prophylaxis	31%	18%-60%	[43,44]
5-year mortality due to variceal bleeding	20%	20%-80%	[36]
Annual incidence of hepatocellular carcinoma	2.5%	-	[38]
3-year mortality due to hepatocellular carcinoma	49.2%	46.4%-77.2%	[35]
Liver biopsy *			[45]
Sensitivity	93%	89%-100%	
Specificity	95%	92%-100%	
Fib-4 index			[25]
Optimal high cutoff (>3.25)			
Sensitivity	38%	35%-41%	
Specificity	97%	95%-100%	
Optimal low cutoff (<1.30)			
Sensitivity	84%	74%-85%	
Specificity	69%	65%-71%	
Indeterminate results	32%	12%-46%	[25]
Magnetic resonance elastography *Optimal cutoff for cirrhosis (4.7)*			[46,47]
Sensitivity	80%	60%-97%	
Specificity	86%	84%-93%	
Failure rates	4%	4%-6%	[33,34]
Vibration controlled transient elastography			[22,32,46,47]
Optimal cutoff for cirrhosis (11.8)			
Sensitivity	80%	78%-95%	
Specificity	81%	85%-89%	
Failure rates	7.1%	3.5%-50%	[22,32]

* Considering optimal liver tissue specimens (≥ 2.5 cm in length and ≥ 10 portal tracts)
Specificity and sensitivity for cirrhosis were calculated using pooled data from published studies.

An institutional cost perspective was adopted, expressed in 2017 U.S. dollars. Direct costs for LB, MRE and VCTE were obtained from Medicare reimbursement data (Appendix A Table A.4).

2.2.4 Analysis

We simulated the costs for a particular strategy, which included the costs of (1) each diagnostic strategy, (2) HCC and EV screenings, (3) periodic clinic visits, (4) management of LB-related complications (i.e., bleeding) and (5) liver transplantation or HCC-related liver resection. We also simulated the diagnostic accuracy for a particular strategy. Diagnostic accuracy was defined as the percentage of correct diagnosis, which equals the sum of numbers of people who were correctly identified with or without cirrhosis (excluding the numbers after the confirmation test), divided by the total population. We next calculated the cost per correct diagnosis, which was used as a proxy in the ensuing cost-effectiveness analysis. We lastly calculated the incremental cost-effectiveness ratios (ICERs) and cost per death prevented (CPDP). ICERs were defined as the incremental cost for each additional correct diagnosis: the strategies were sorted by ascending order of cost, and the current least costly strategy was compared with the previous least costly strategy. The formula to calculate ICERs is shown in the supplemental material. If the accuracy of the current least costly strategy was lower than the previous one, then the current strategy was considered dominated. CPDP was calculated using the same logic, and we used mortality instead of diagnostic accuracy.

We further plotted per-patient total cost and diagnostic accuracy, and per-patient total cost and mortality for each strategy. Any strategy, such that no other strategies would yield lower cost and higher accuracy (lower mortality) simultaneously, was considered to be a dominant strategy. In the presence of more than one dominant strategy, the dominant strategies formed an efficiency frontier; strategies below and to the right of the frontier were considered to be dominated.

To assess the robustness of model results, we performed sensitivity analyses on cirrhosis prevalence (0%-12%). In addition, we performed one-way sensitivity analyses on test characteristics and costs with cirrhosis prevalence being fixed at 0.27%.

2.3 Results

2.3.1 Base case.

Low prevalence of cirrhosis (0.27%). Fib-4 alone correctly classifies the lowest percentage of persons (57%) and is the reference strategy for the change in percentage of persons correctly classified, while Fib-4+LB correctly classified 97.7%, the highest percentage (Table 2.2). Figure 2.4 shows diagnostic accuracy by cost per person for each strategy. The solid line represents the efficiency frontier, which identifies strategies with the lowest cost and highest accuracy, and includes: Fib-4+VCTE (89.3%, $401), Fib-4+MRE (92.4%, $491) and Fib-4+LB (97.7%, $729). Relative to Fib-4+VCTE, which is the least costly strategy with 89.3% accuracy, ICERs for the frontier strategies range from $2,864 per additional correct diagnosis for Fib-4+MRE to $4,454 for Fib-4+LB (Table 2.2). Figure 2.4 and Appendix A Table A.5 display mortality by cost per person for each strategy. Compared to do-nothing strategy, which yields highest mortality (46 deaths) and lowest cost ($10), Fib-4+VCTE (41 deaths, $401), VCTE+LB (39 deaths, $613), VCTE (38 deaths, $901) and LB (36 deaths, $1663) show the best combination of cost versus mortality, respectively.

Appendix A Tables A.6-A.14 depict detailed results on diagnostic accuracy for all strategies.

Table 2.2. Accuracy and cost-effectiveness of different diagnostic strategies. A microsimulation analysis based on 100,000 NAFLD patients considering a cirrhosis prevalence of 0.27%.

Strategy	Number of correctly identified out of 270 people with cirrhosis	Number of correctly excluded out of 99730 people without cirrhosis	Percentage of people correctly classified	Change in percentage of people correctly classified*	Cost per person ($)	Cost per correct diagnosis ($)	Mortality (bleeding)	Mortality (HCC)	Total mortality cases	ICER
No test	N/A	N/A	N/A	N/A	10	N/A	25	5	46	N/A**
Fib-4+VCTE	116	89,145	89.3%	32.2%	401	450	21	5	42	**Least costly**
Fib-4+MRE	120	92,272	92.4%	35.4%	491	531	22	5	42	2,864
VCTE+LB	187	91,769	92.0%	34.9%	612	667	19	4	40	4,454
Fib-4+LB	145	97,592	97.7%	40.7%	729	747	21	5	42	Dominated
MRE+LB	193	95,071	95.3%	38.2%	888	932	19	4	40	Dominated
VCTE	201	75,046	75.2%	18.2%	900	1,197	18	4	39	Dominated
Fib-4	70	56,966	57.0%	**Reference**	908	1,592	21	5	41	Dominated
MRE	207	82,337	82.5%	25.5%	1,109	1,344	19	4	39	Dominated
LB	251	94,744	95.0%	38.0%	1,663	1,751	17	4	37	Dominated

Abbreviations: VCTE, vibration controlled transient elastography; **LB,** liver biopsy; **MRE,** Magnetic resonance elastography; **HCC,** hepatocellular carcinoma; **ICER,** incremental cost effectiveness ratio.

Total mortality includes mortality cases from variceal bleeding, HCC and other than bleeding or HCC.

Strategies are listed in order of increasing costs.

* It represents difference between tests and the reference strategy (Fib-4).

** No ICER value for do-nothing strategy, since there is no accuracy

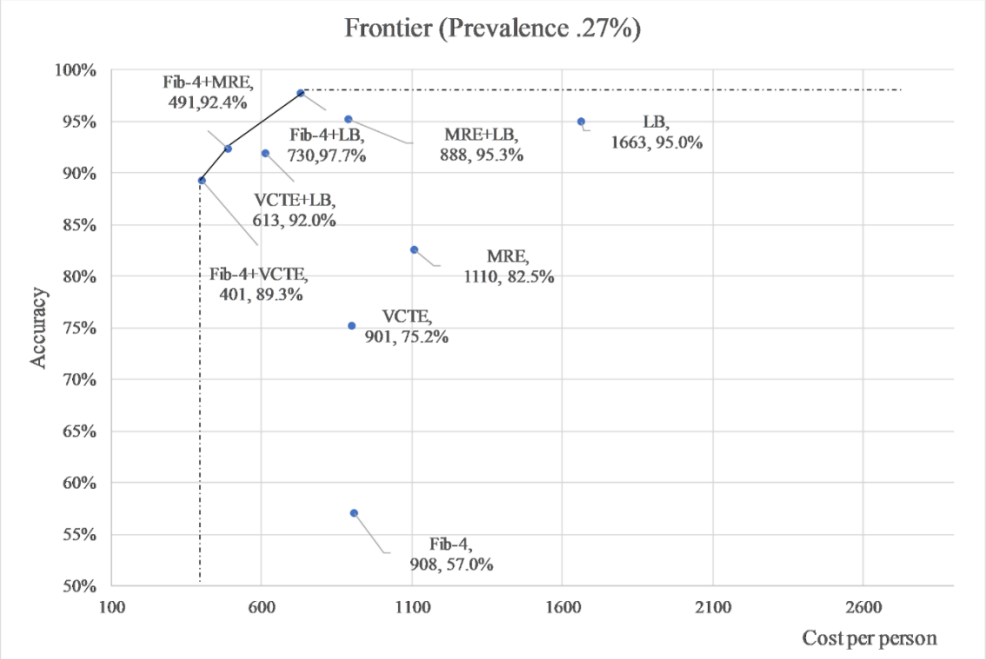


Figure 2.4. Incremental cost-effectiveness "Frontier" of 9 diagnostic strategies: Cost per person vs accuracy considering population-based prevalence (0.27%).

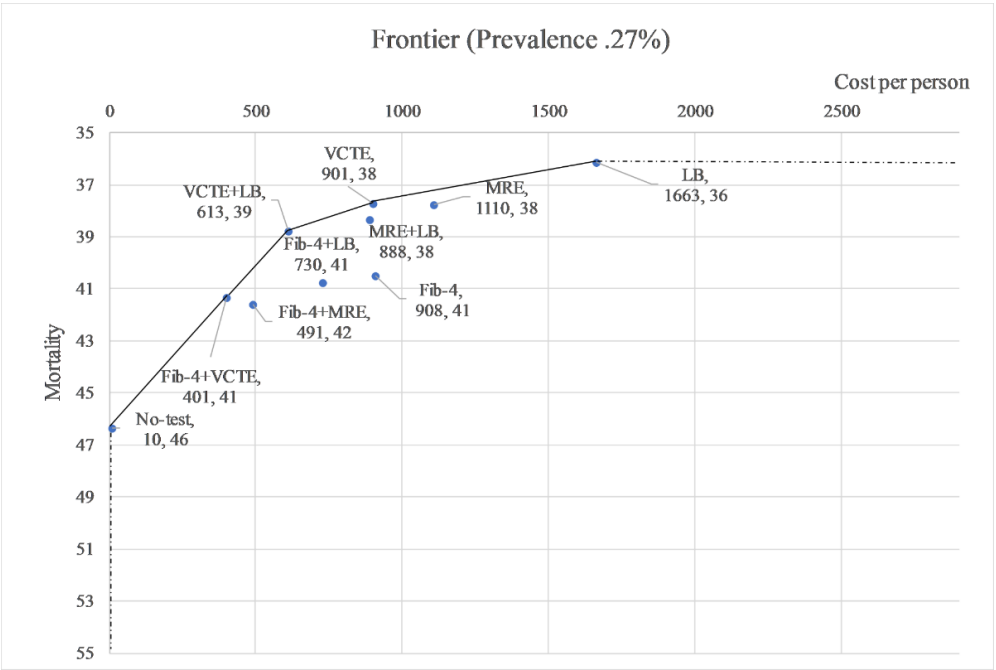


Figure 2.5. Incremental cost-effectiveness "Frontier" of 9 diagnostic strategies: Cost per person vs mortality considering population-based prevalence (0.27%).

The points on the line are the strategies on the frontier and considered dominating strategies, and points below the line are the less cost-effective strategies.

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography.

Intermediate prevalence of cirrhosis (2%). Based on case results for intermediate prevalence of cirrhosis, both the cost per person for each strategy and the cost per correct diagnosis per person are higher for an intermediate prevalence of cirrhosis as compared with low cirrhosis prevalence. Fib-4 alone remains the least accurate strategy (56.5%). Base results for intermediate cirrhosis prevalence closely parallel results for low prevalence with the same three least costly strategies which are Fib-4+VCTE (88.5% accuracy, cost of $690), Fib-4+MRE (91.6%, $781), and Fib-4+LB (97%, $1,060) (Table 2.3). VCTE+LB has the same cost than Fib-4+LB, but it yields lower diagnostic accuracy (Table 2.3 and Figure 2.5). Figure 2.3 shows 3 strategies on the efficiency frontier, including the three same frontier strategies for the low prevalence case, with ICERs ranging from $2,918 (Fib-4+MRE) to $5,156 (Fib-4+LB) per additional correct diagnosis (Table 2.3). Results based on cost versus mortality analysis (Figure 2.7 and Appendix A Table A.15), show the same four strategies on the frontier (Fib-4+VCTE [307, $690], VCTE+LB [285, $1,060], VCTE [279, $1,347] and LB [270, $2,183]) than those results based on low cirrhosis prevalence.

Table 2.3. Accuracy and cost-effectiveness of different diagnostic strategies. A microsimulation analysis based on 100,000 NAFLD patients considering a cirrhosis prevalence of 2%.

Strategy	Number of correctly identified out of 2000 people with cirrhosis	Number of correctly excluded out of 98000 people without cirrhosis	Percentage of people correctly classified	Change in percentage of people correctly classified*	Cost per person ($)	Cost per correct diagnosis ($)	Mortality (bleeding)	Mortality (HCC)	Total mortality cases	ICER
No test	N/A	N/A	N/A	N/A	77	N/A	187	41	344	N/A**
Fib-4+VCTE	860	87,599	88.5%	32.0%	690	780	158	30	307	**Least costly**
Fib-4+MRE	888	90,671	91.6%	35.1%	781	853	159	30	307	2,918
VCTE+LB	1,076	95,899	97.0%	40.5%	1,060	1,093	155	29	302	5,156
Fib-4+LB	1,382	90,177	91.6%	35.1%	1,060	1,158	142	23	285	Dominated
MRE+LB	517	55,978	56.5%	**Reference**	1,236	2,187	153	28	300	Dominated
VCTE	1,428	93,421	94.8%	38.4%	1,329	1,401	142	24	285	Dominated
Fib-4	1,486	73,744	75.2%	18.7%	1,347	1,791	138	22	279	Dominated
MRE	1,536	80,909	82.4%	26.0%	1,557	1,889	139	22	280	Dominated
LB	1,860	93,100	95.0%	38.5%	2,183	2,299	131	19	270	Dominated

Abbreviations: VCTE, vibration controlled transient elastography; **LB,** liver biopsy; **MRE,** Magnetic resonance elastography; **HCC,** hepatocellular carcinoma; **ICER,** incremental cost effectiveness ratio.

Total mortality includes mortality cases from variceal bleeding, HCC and other than bleeding or HCC.

Strategies are listed in order of increasing costs.

* It represents difference between tests and the reference strategy (Fib-4).

** No ICER value for do-nothing strategy, since there is no accuracy

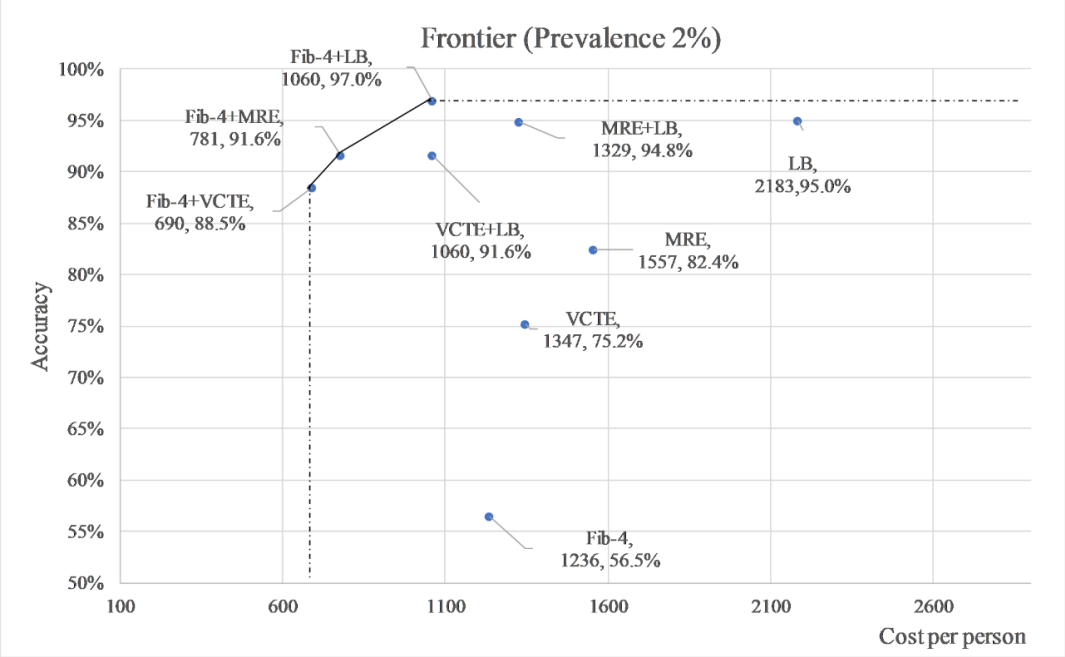


Figure 2.6. Incremental cost-effectiveness "Frontier" of 9 diagnostic strategies: Cost per person vs accuracy considering population-based prevalence (2%).

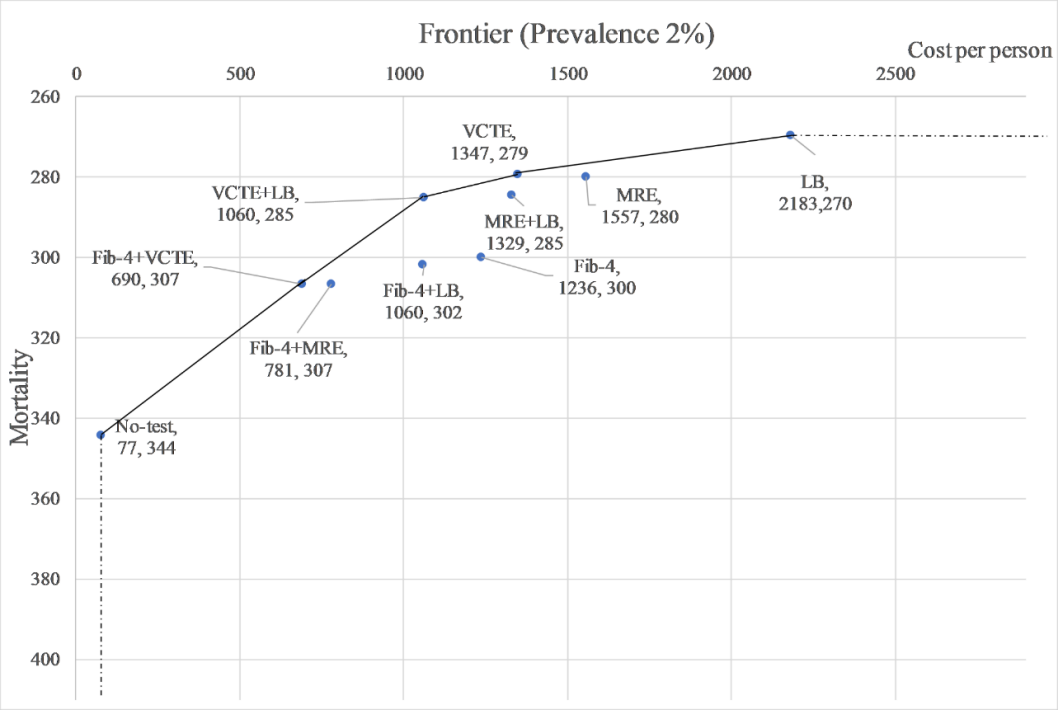


Figure 2.7. Incremental cost-effectiveness "Frontier" of 9 diagnostic strategies: Cost per person vs mortality considering population-based prevalence (2%).

The points on the line are the strategies on the frontier and considered dominating strategies, and points below the line are the less cost-effective strategies.

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography.

High prevalence of cirrhosis (4%). Baseline results for the high prevalence of cirrhosis closely parallel those from low and intermediate prevalence (Table 2.4, Figure 2.6). The same three strategies (Fib-4 followed by either VCTE, MRE or LB) comprise the 3 least costly strategies, with accuracies ranging from 87.5% for Fib-4+VCTE to 96.1% for Fib-4+LB. As expected, due to a higher proportion of cirrhosis, both costs per person and cost per correct diagnosis are higher than in the two previous prevalence scenarios, ranging respectively from $1,024 for Fib-4+VCTE to $1,441 for Fib-4+LB and from $1,170 for Fib-4+VCTE to $1,500 for Fib-4+LB. Fib-4 alone remains the least accurate strategy (55.9%). Similar to findings under conditions of intermediate prevalence, the same three frontier strategies appear on the efficiency frontier, with ICERs ranging from $2,921 to $5,956 per additional correctly diagnosed case. Cost versus mortality analysis also show the same four diagnostic strategies on the frontier, with Fib-4+VCTE displaying highest mortality (613) followed by VCTE+LB (568), VCTE (559) and LB (540) respectively (Figure 2.9 and Appendix A Table A.16).

.

Table 2.4. Accuracy and cost-effectiveness of different diagnostic strategies. A microsimulation analysis based on 100,000 NAFLD patients considering a cirrhosis prevalence of 2%.

Strategy	Number of correctly identified out of 4000 people with cirrhosis	Number of correctly excluded out of 96000 people without cirrhosis	Percentage of people correctly classified	Change in percentage of people correctly classified*	Cost per person ($)	Cost per correct diagnosis ($)	Mortality (bleeding)	Mortality (HCC)	Total mortality cases	ICER
No test	N/A	N/A	N/A	N/A	157	N/A	373	81	686	N/A**
Fib-4+VCTE	1872	85659	87.5%	31.7%	1024	1170	316	60	613	**Least costly**
Fib-4+MRE	1863	88735	90.6%	34.7%	1114	1230	318	60	613	2,921
VCTE+LB	2152	93942	96.1%	40.2%	1441	1500	310	57	603	5,956
Fib-4+LB	3029	88073	91.1%	35.2%	1579	1733	283	47	568	Dominated
MRE+LB	2224	53645	55.9%	**Reference**	1616	2892	306	56	599	Dominated
VCTE	3006	91366	94.4%	38.5%	1840	1950	284	47	569	Dominated
Fib-4	3237	71975	75.2%	19.3%	1861	2474	276	44	559	Dominated
MRE	3221	79109	82.3%	26.5%	2077	2523	277	44	559	Dominated
LB	3720	91200	94.9%	39.1%	2777	2925	262	39	540	Dominated

Abbreviations: VCTE, vibration controlled transient elastography; **LB,** liver biopsy; **MRE,** Magnetic resonance elastography; **HCC,** hepatocellular carcinoma; **ICER,** incremental cost effectiveness ratio.

Total mortality includes mortality cases from variceal bleeding, HCC and other than bleeding or HCC.

Strategies are listed in order of increasing costs.

* It represents difference between tests and the reference strategy (Fib-4).

** No ICER value for do-nothing strategy, since there is no accuracy

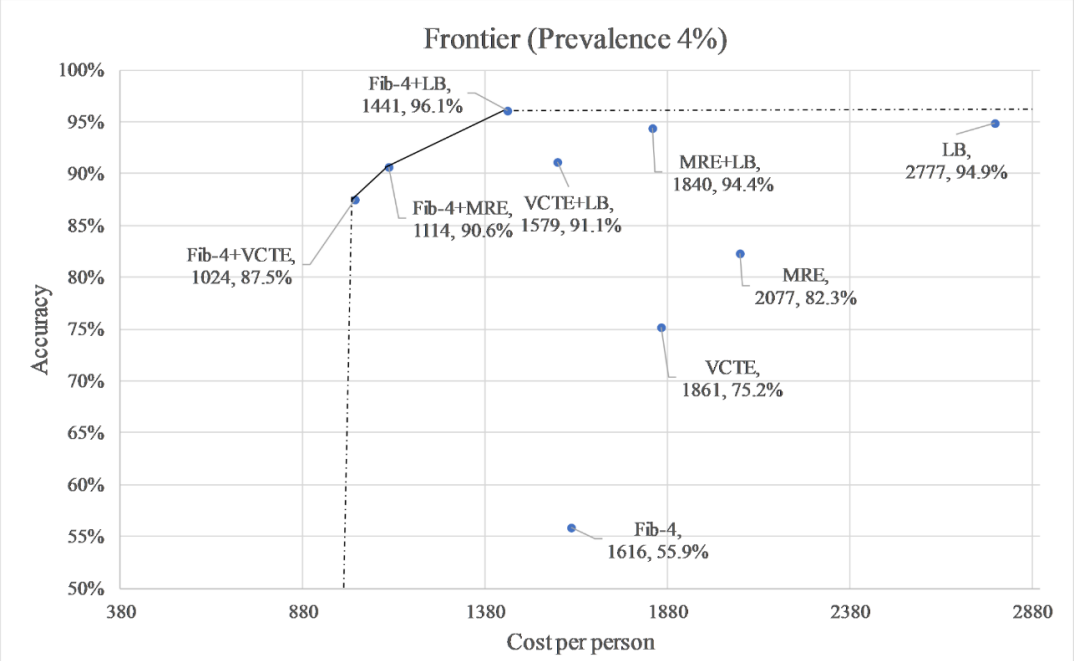


Figure 2.8. Incremental cost-effectiveness "Frontier" of 9 diagnostic strategies: Cost per person vs accuracy considering population-based prevalence (4%).

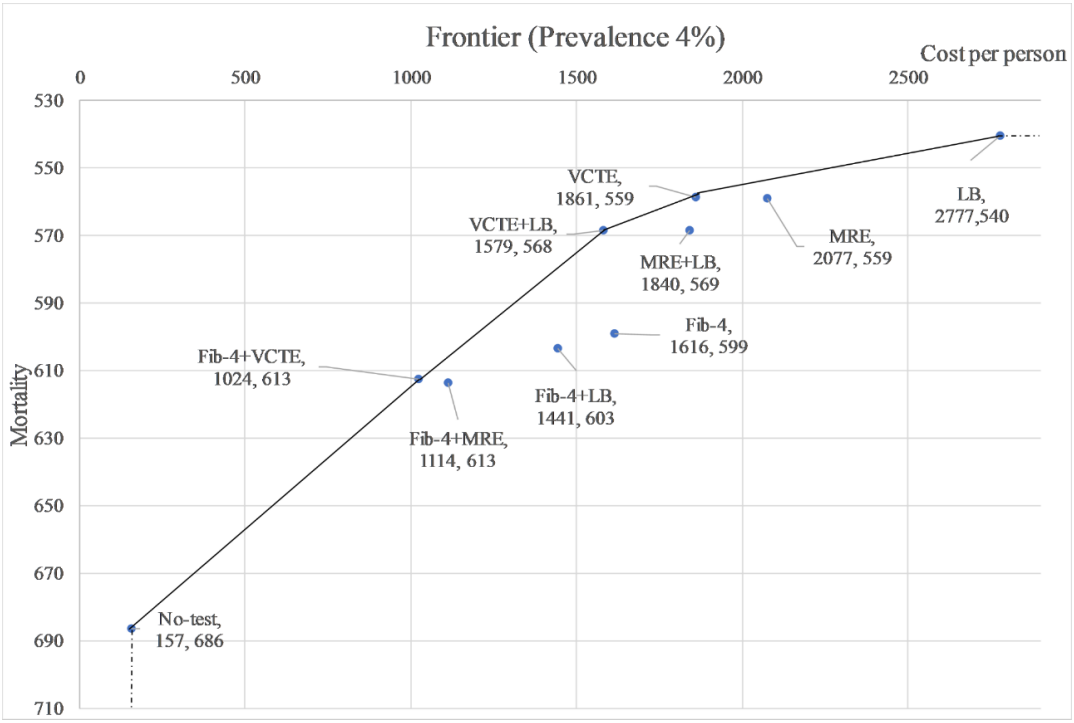


Figure 2.9. Incremental cost-effectiveness "Frontier" of 9 diagnostic strategies: Cost per person vs mortality considering population-based prevalence (4%).

The points on the line are the strategies on the frontier and considered dominating strategies, and points below the line are the less cost-effective strategies.

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography.

2.3.2 Sensitivity analyses.

Prevalence of cirrhosis. When the prevalence of cirrhosis is extended to as high as 12%, cost per correct diagnosis increases for all strategies, as shown in Appendix A Figure A.2. The largest cost increases are observed for Fib-4, VCTE, and MRE, with respective increases of $4,276, $4,030, and $3,732 from the previously considered low prevalence (0.27%). Cost increases are the lowest for the two non-invasive combination test strategies of FIB-4+VCTE and FIB-4+MRE, with respective increases of $2,364 and $2,298 from the previously considered low prevalence scenario (0.27%).

Test characteristics. Results of one-way sensitivity analyses on each test's sensitivity and specificity are shown in Appendix A Figure A.3A-D. Within the ranges of test characteristics, the general pattern of cost per correct diagnosis are comparable: negligible effects of sensitivity on either individual or combination strategies, negligible effects of specificity for combination strategies, and a modest cost reduction for individual tests as specificity increases.

Test costs. Results of 8 one-way sensitivity analyses on the cost of each test are shown in Appendix A Table A.17 and in Appendix A Figure A.4. Note that the baseline values were set at Medicare average price Appendix A Table A.4, i.e., $0, $1,411, $150.34 and $544.18 for Fib-4 LB, VCTE and MRE. For the one-way sensitivity analysis on each cost parameter, alternative cost values are based on the national average for facility fees among hospital-based clinics and the 90th percentile for charges submitted by hospital-based clinics; see Appendix A Table A.17.

Considering sensitivity in the tradeoff of cost vs accuracy with respect to the cost parameters, Appendix A Table A.18 shows the percentage of times (9 in total) each strategy is on efficiency frontier. Fib-4+VCTE, Fib-4+MRE, and Fib-4+LB are on efficiency frontier in 89% of the scenarios, respectively. In contrast, MRE+LB and VCTE+LB are on the frontier only when the costs of MRE and LB are high, and in 11% of the scenarios, respectively. The Tornado plot of cost per correct diagnosis shows Fib-4+VCTE to be the least costly strategy with a range of $397 to $590, whereas LB and MRE, each alone, are the two most costly strategies (Appendix A Figure A.4).

2.4 Discussion

Since cirrhosis is the major determinant of long-term morbidity and mortality in patients with NAFLD, there is a critical need to detect cirrhosis before complications occur, which are associated with a high mortality and increased healthcare utilization. One strategy to alleviate the burden of liver disease is to identify at-risk asymptomatic individuals and offer them screening via non-invasive diagnostic approaches. The sequential combination of two NITs or a NIT test plus liver biopsy may detect cirrhosis more accurately,[28,29] however, whether these approaches are cost-effective is unknown. We developed a decision model to compare accuracy and cost impact of different single or sequential tests for cirrhosis diagnosis in individuals with NAFLD. This study suggests that, among all diagnostic strategies, the use of Fib-4 followed by VCTE is likely the most cost-effective strategy for screening or detecting cirrhosis among patients with NAFLD either in the setting of general population, or specialty clinics and tertiary referral centers, respectively. Other diagnostic strategies such as Fib-4 followed by either MRE or LB, and VCTE followed by LB are likely cost-effective approaches but might have higher costs than Fib-4 followed by VCTE across all cirrhosis prevalence scenarios.

There is a current knowledge gap regarding the cost-effectiveness of single versus sequential combined screening strategies for cirrhosis in NAFLD. Our findings support previous work suggesting that the use of sequential combination tests is more cost-effective than are single tests.[28,29,48] Combination of tests yields higher accuracy than single tests resulting in higher proportion of patients correctly diagnosed with cirrhosis and subsequent reduction in costs from misdiagnosis. Recently, Majumdar et al. explored the clinical consequences of using the minimum acceptable diagnostic accuracy of NITs in diagnosing cirrhosis with equivalence to LB in terms of mortality.[48] They found that sequential NITs performed better than single NITs at different predefined prevalence of cirrhosis.

Among the 5 combination tests compared in the base case analysis, Fib-4+VCTE ranks first due to the best combination of cost per correct diagnosis and diagnostic accuracy. Fib-4+MRE ranks second, with higher percentage of people correctly classified (ranging from 90.6% to 92.4% vs. 87.5%-89.3% with Fib-4+VCTE) but higher costs. The combinations of Fib-4 with LB or VCTE with LB ranked third and fourth, respectively, across all cirrhosis prevalence settings. Although

both combinations yield higher accuracy, they have higher costs per correct diagnosis as compared with either FIB-4+VCTE or Fib-4+MRE. Finally, the combination of MRE with LB ranked fifth among all combination tests.

Overall, using MRE together with either LB or Fib-4 moderately increases overall costs and therefore reduces cost-effectiveness as compared to combinations that include VCTE plus Fib-4 or LB. The results highlight the importance of diagnostic accuracy, with costs being more influential on strategies including MRE. This finding suggests that, due to its point-of-care availability and more affordable cost, a VCTE based strategy may be the most attractive approach when considering potential population-based screening program as well as in the setting of specialist clinics.

If the goal is to avoid liver biopsy, Fib-4+VCTE is a very cost-effective strategy among NITs. This strategy could be particularly important in the community setting or in resource-limited areas where Fib-4+VCTE can be used to screen for cirrhosis among patients with NAFLD. Use of NITs by primary care providers may reduce unnecessary or late specialty referrals which are associated with either overuse of health care services or with suboptimal patient outcomes respectively. In the setting of higher prevalence of cirrhosis such as referral centers, either Fib-4 together with VCTE or MRE may be useful tools for detecting cirrhosis.

This analysis evaluated the cost-effectiveness of a comprehensive list of screening strategies for cirrhosis diagnosis in the context of NAFLD with varying prevalence of cirrhosis. Other strengths of this study include: (1) using liver biopsy as the reference standard for cirrhosis diagnosis and blood- and imaging-based tests, (2) simulated distribution of outcomes and its treatments considering real-world data, and (3) assessment of uncertainty including a wide range of cirrhosis prevalence, sensitivity and specificity values, as well as assuring the statistical confidence in comparative studies by running the microsimulation with a sufficiently large cohort multiple times for each strategy.

This analysis has limitations as well, including: (1) results are based on the U.S. costs and may not generalize to other healthcare systems; (2) MRI costs were used as a proxy for MRE; MRE is a

relatively new diagnostic imaging technology without a CPT code for insurance reimbursement currently, although it is being used often in the United States for routine clinical care; (3) since the sensitivity and specificity of non-invasive tests in the context of a primary care setting are not available, this information was extrapolated from studies conducted at tertiary referral centers; and (4) finally, our cost-effectiveness is not a traditional one, i.e., cost per life-year save or cost per quality adjusted life year. We used cost per correct diagnosis as a proxy for cost-effectiveness for several reasons. First, the time horizon for this analysis is short such that the risk of hard outcomes such as mortality due to hepatic and cardiovascular events is unlikely. Second, not well-described are the natural history of cirrhosis due to NAFLD and the effects of identifying and treating large esophageal varices and hepatocellular carcinoma in this setting.

In conclusion, this study suggests that Fib-4 followed by either VCTE, MRE or LB are cost-effective strategies for identifying cirrhosis in populations where the prevalence of cirrhosis varies between 0.27%-4%. Compared to the combination of FIB-4 and VCTE, the ICERs were higher for the combination of FIB-4 and MRE were lower than for the combination of FIB-4 and liver biopsy. If the goal is to avoid liver biopsy, the combination of Fib-4+VCTE with its lower costs and accessibility is likely the preferred strategy for the screening of cirrhosis in the setting of general or community-based populations.

3. ESSAY 2: THE HEALTH AND ECONOMIC IMPACT OF USING A SUGAR SWEETENED BEVERAGE TAX TO FUND FRUIT AND VEGETABLE SUBSIDIES IN NEW YORK CITY: A MODELING STUDY

3.1 Introduction

Healthful diets are an important protective factor for cardiovascular disease (CVD).[49,50] Specifically, low fruit and vegetable (FV) intake and high sugar-sweetened beverage (SSB) consumption are independently associated with an increased risk of developing CVD.[51–54] Yet, approximately half of US adults report consuming at least one SSB on a given day.[55] Moreover, only one in three women and one in five men report eating the recommended amount of fruits and vegetables per day (5 +servings/day) nationally.[56]

As CVD continues to place a heavy burden on society and the healthcare system in the USA, public health professionals and policymakers have increasingly looked to SSB taxes as a strategy for decreasing sugar intake and lowering the risk of CVD in the population. To date, 8 jurisdictions enacted SSB excise taxes across the USA—including in Albany (NY), Berkeley (CA), Boulder (CO), Oakland (CA), Philadelphia (PA), San Francisco (CA), Santa Fe (NM), Seattle (WA), and Cook County (IL).[57] Evidence indicates that the taxes discouraged the purchase and consumption of SSBs.[58–60] An SSB excise tax also generates new revenue that can be earmarked, or directed, towards programs to promote health. One such program that has been implemented in several cities with tax revenues is FV financial incentives for those shopping with Supplemental Nutrition Assistance Program (SNAP) benefits,[61] which have been linked to an increase in FV purchases in farmers markets,[62] mobile produce markets,[63] and supermarkets[64,65]; and have the potential to increase FV intake that is protective against CVD.[66] Pairing SSB taxes and FV subsidies may improve dietary behaviors from multiple dimensions and have a greater positive combined effect on health than singly policy action alone.[67]

In New York City (NYC), the high cost of living and the existence of areas that are and have been historically impacted by food apartheid (areas with limited access to fresh FVs) make fresh FVs less affordable and accessible.[68,69] These and multiple other factors may contribute to the low observed prevalence of consumption of 5 servings of FVs amongst NYC adults of 10%; nationally, this percentage is 12%.[68,70] In addition, SSB consumption in NYC remains high and has plateaued after several years of decline, with 24% of residents reporting drinking one or more SSB per day, and 84% reporting drinking an SSB in a typical week.[71] A potential implementation of an SSB tax and an FV subsidy program could result in substantial public health and economic benefits.

In this study, we used a microsimulation model of CVD to assess the impact of implementing SSB taxes and FV subsidies on long-term CVD outcomes and healthcare costs in NYC. We simulated the potential implementation of each policy alone as well as funding FV subsidies with an SSB tax (a combined policy). We also assessed the cost-effectiveness of each policy in preventing CVD compared to the status quo.

3.2 Methods

3.2.1 Model Development

We developed a microsimulation model of CVD for NYC adults based on the well-established CVD Policy Model.[72–75] Model details and a model schematic are included in Appendix B. Briefly, the model simulates healthy individuals (i.e., no history of CVD) and their risk of CVD over time. Within each year of the simulation, individuals are at risk of experiencing coronary heart disease (CHD), stroke, both CHD and stroke, CVD-related death, and non-CVD-related death (Fig. 1). The annual probability of incident CHD, incident stroke, and non-CVD related death are estimated by functions accounting for age, body mass index (BMI), smoking status, systolic blood pressure, diabetes status, high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and estimated glomerular filtration rate (a list of model parameters is presented in Appendix B Table B.1). Once a CHD or stroke event occurs, individuals are at risk of secondary or recurrent CVD events and CVD-related death (Appendix B Table B.2). We assumed each individual can experience at most two CVD-related events per year. As individuals progress through the model,

their CVD event history, survival, quality-adjusted life years (QALYs), direct healthcare costs, and program implementation costs are recorded at 10 years, 20 years, 40 years, and over their entire lifetime (death or 100 years of age). The model was programmed in R (version 3.6.1). The study was approved by the Institutional Review Board at the Icahn School of Medicine at Mount Sinai.

3.2.2 Simulated Population

We simulated a population representative of NYC adults by sampling 10,000 individuals and their characteristics from the NYC Health and Nutrition Examination Survey (NYC HANES, 2003–2004 and 2013–2014). We matched NYC HANES participants on CVD risk factors to participants from the National Heart, Lung, and Blood Institute Pooled Cohorts for whom lifetime CVD risk factor trajectories were previously developed (Appendix B Table B.3).[76–78] We assigned each individual daily intake of SSB and FV usingdata from the 2018 NYC Community Health Survey.[79] We fit truncated normal distributions using race and sex stratified mean per-capita intake of FVs (number of cups) and SSBs (number of drinks) from daily dietary recall questions (Appendix B Table B4).

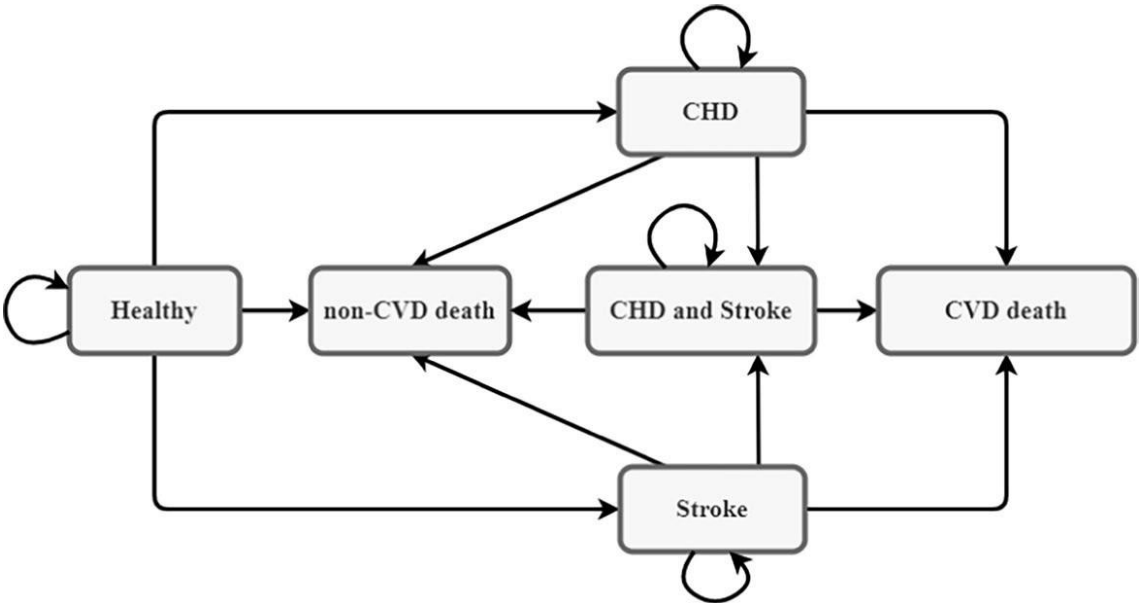


Figure 3.1Model schematic. Notes: CHD, coronary heart disease; CVD, cardiovascular disease

3.2.3 Policy Scenarios

We modeled three policy scenarios, including (1) an SSB tax; (2) an FV subsidy; and (3) a combined policy that uses the SSB tax revenue to fund FV subsidies. We compared each of the above policy scenarios to the status quo. We modeled the effect of the SSB tax and FV subsidy on consumption outcomes through price changes and their associated price elasticity based on Equation 3.1 below (Appendix B Table B.2). To model the relationship, we assumed that (1) the effects from an SSB tax and an FV subsidy were independent (i.e.,an SSB tax only influences SSB consumption and FV subsidies would only impact FV consumption), (2) the time lag between policy implementation and changes in SSB and FV consumptions is less thana year, and (3) the policy effects remain constant as long as taxes and subsidies continue.[80]

3.2.4 Change in SSB and F&V consumption

$$C_{new} = C_{old} \left(\frac{p_{new}}{p_{old}}\right)^{\varepsilon} \qquad \text{Equation 3.1}$$

where C_{old} and C_{new} are SSB consumption and FV consumption before and after policy implementation, p_{old} and p_{new} are SSB and FV prices before and after policy implementation, and

ε is price elasticity, which reflects the percentage of consumption change with a 1-percent price change. We used the SSB price change percentage and mean FV prices to derive the purchase price percentage changes for FV subsidies (Appendix B Table B.5). We assumed tax revenues would be equally distributed to fund FV subsidies.

3.2.5 The Effects of SSB and FV Consumption Changes on CVD and Diabetes Risk

We estimated the relative risk (RR) of CHD and stroke incidence with SSB and FV consumption based on findings from recent meta-analyses (Appendix B Table B.6).[81,82] We also included an RR for the effect of SSB and FV intake on the incidence of diabetes mellitus.[50,82] We modeled the effects of changes in consumption on the risk of CHD, stroke, and diabetes as $RR = RR_{incremental}{}^{(C_{new}-C_{old})}$, where $RR_{incremental}$ is the medical costs plus policy and implementation costs. RR per one-unit reduction in SSB or FV consumption per day and is multiplied by the probability of an incident event without SSB or FV policies.

3.2.6 Cost and Utility Model Parameters

We estimated the implementation costs for SSB taxes to be 2% of the tax revenue collected and for FV subsidies to be 20% of the subsidies in the first year and 5% for the years afterward. Details about our estimation of the policy implementation costs are provided in Appendix B. We estimated the QALYs for different disease stages based on the published literature.[83–85] The model includes disutility associated with acute CHD and stroke events that are applied for 30 days and subsequent chronic disutility that is applied each cycle afterward. The healthcare costs and disutility values are presented in Tables B.7-B.8 in Appendix B. The QALY and cost parameters were also used in the other studies with the microsimulation version of the CVD Policy Model.[72,73] Healthcare costs were inflated to 2019 US dollars using the medical component of the US Consumer Price Index.

3.2.7 Model Validation

Our model was calibrated to match contemporary CHD, stroke, and mortality rates for the US from the Centers for Disease Control and Prevention (CDC), National Hospital Discharge Survey, National Inpatient Sample, National Vital Statistics System, and NHLBI Pooled Cohort Study, and cross-validated against the dynamic population version of the CVD Policy Model (Figure B.6-B.8

in Appendix B). We then compared the estimated CVD, non-CVD, and all-cause mortality rates from our model to those observed in the NYC metropolitan area derived from the CDC's Wide-ranging ONline Data for Epidemiologic Research (WONDER).[86]

3.2.8 Statistical Analyses

For the primary analysis, we compared the costs, QALYs, and cost-effectiveness of the status quo, SSB taxes only, FV subsidies only, and a combined policy over 10 years. Each policy was evaluated from the healthcare sector perspective (direct medical costs regardless of payer) and societal perspective (direct medical costs plus policy and implementation costs). Incremental cost-effectiveness ratios (ICERs) for each policy were calculated as the mean difference in costs divided by the mean difference in QALYs. Future costs and QALYs were discounted annually at 3%.[87] We used a threshold of $50,000/QALY to determine if a strategy was cost-effective.[88] We also used the incremental net monetary benefit (INMB) to determine the cost-effectiveness of a strategy. INMB represents the monetary value of a strategy at a given willingness-to-pay threshold and is calculated as: INMB = incremental QALYs * willingness-to-pay – incremental costs. For this analysis, when the monetized incremental health gains are greater than the incremental costs (i.e., INMB>$0), the strategy is cost-effective relative to the status quo. Further, the strategy with the highest INMB is the most cost-effective and, thus, the preferred strategy at the willingness-to-pay threshold.

We conducted deterministic sensitivity analyses to assess the potential impact of time horizon, price elasticity, and policy implementation costs on cost-effectiveness results. We accounted for joint uncertainty of model parameters in the cost-effectiveness analysis by probabilistically sampling parameter values from prespecified distributions in 1000 model iterations (Appendix B Table B.5). Results are presented as the mean and 95% uncertainty intervals (UI; 2.5th to 97.5th percentile) of 1000 iterations. Our analysis adhered to the requirements from the Consolidated Health Economic Evaluation Reporting Standards (CHEERS) 2022 Checklist (Appendix B Table B.9).

3.3 Results

Compared with the status quo over 10 years, ourmodel projected that, per 10,000 adults in NYC, an SSB tax would prevent 26 (95%UI, 4 to 92) CVD events, FV subsidies 16 (95%UI, − 1 to 36) CVDevents, and a combined policy 41 (95%UI, 15 to 73) CVD events (Table 3.1). Compared with the status quo, an SSB tax would increase total QALYs by 24 (95%UI, 4 to 51), FV subsidies by 15 (95%UI, 2 to 35), and the combined policy by 37 (95%UI, 9 to 72).

Table 3.1. Projected health and economic outcomes in 10 years under different policies

	Status quo	SSB taxes	FV subsidies	SSB taxes + FV subsidies
Healthcare Outcomes				
CHD events	592 (542 to 648)	574 (523 to 631)	584 (534 to 638)	566 (518 to 622)
Stroke events	275 (241 to 313)	267 (232 to 304)	268 (233 to 304)	260 (225 to 296)
CVD deaths	165 (133 to 202)	161 (128 to 196)	163 (129 to 199)	158 (125 to 193)
QALYs	81,702 (81,380 to 81,996)	81,726 (81,406 to 82,020)	81,717 (81,400 to 82,020)	81,739 (81,416 to 82,036)
Costs (2019 USD, thousands)				
Total (Societal)	670,900 (661,111 to 680,489)	666,869 (657,206 to 676,306)	674,380 (664,683 to 683,884)	670,366 (660,656 to 679,850)
Healthcare	670,900 (661,111 to 680,489)	670,153 (660,544 to 679,564)	670,434 (660,745 to 679,937)	669,703 (659,982 to 679,049)
Prevented Healthcare Outcomes				
CHD events	-	18 (3 to 75)	9 (-2 to 20)	26 (9 to 46)
Stroke events	-	8 (1 to 17)	7 (1 to 16)	15 (6 to 27)
CVD deaths	-	4 (-2 to 151)	3 (-3 to 10)	8 (-3 to 20)
QALYs gained	-	24 (4 to 51)	15 (2 to 35)	37 (9 to 72)
Incremental Costs (2019 USD, thousands)				
Total (Societal)	-	-4,030 (-4,580 to -3,640)	3,480 (3,130 to 3,770)	-540 (-1,150 to -20)
Healthcare	-	-750 (-1,390 to -260)	-470 (-820 to -180)	-1,200 (-1,900 to -620)
Policy	-	-3,280 (-3,530 to -3,040)	3,950 (3,870 to 4,020)	660 (400 to 910)
Implementation	-	70 (60 to 70)	250 (250 to 260)	320 (310 to 330)
Incremental cost-effectiveness				
Healthcare sector perspective				
ICER ($/QALY)	-	Dominated*	Dominated*	Dominant**
INMB ($, thousands)		1,933 (543 to 3,813)	1,210 (382 to 2,467)	3,065 (1,248 to 5,290)
Societal perspective				
ICER ($/QALY)	-	Dominant**	Dominated***	Dominant**
INMB ($, thousands)		5,216 (3,986 to 6,941)	-2,736 (-3,566 to -1,513)	2,402 (682 to 4,492)

Notes: SSB=sugar sweetened beverage; FV=fruit and vegetable; CHD=coronary heart disease; CVD=cardiovascular disease; QALY=quality-adjusted life year; ICER=incremental cost-effectiveness ratio; INMB=Incremental net monetary benefit

[1] All numerical results are presented as mean estimate with (95% UIs)

[2] INMB represents the monetary value of an intervention at a given willingness-to-pay threshold. INMB is calculated as: INMB = incremental QALYs*willingness-to-pay – incremental costs. For this analysis, willingness-to-pay was set at $50,000 per QALY. When the monetized incremental health gains are greater than the incremental costs (i.e., INMB >$0), the strategy is cost-effective relative to the status quo. Further, the strategy with the highest INMB is the most cost-effective and, thus, preferred strategy at the willingness-to-pay threshold.

*Dominated (i.e., costs more and less effective) by combining SSB taxation with F&V subsidies

**Dominant (i.e., cost less and more effective) vs. status quo

***Dominated by SSB taxation alone and combining SSB taxation with F&V subsidies

Each policy was estimated to result in reduced health-care costs compared with the status quo, with the largest savings projected when funding FV subsidies with an SSB tax at $1,200,000 (95%UI, $620,000 to $1,900,000). An SSB tax policy was revenue-generating compared with the status quo and was projected to save $3,280,000 (95%UI, $3,040,000 to $3,530,000) in policy costs over 10 years. However, FV subsidies would increase policy costs by $3,950,000 (95%UI, $3,870,000 to $4,020,000), and the combined policy would increase costs by $660,000 (95%UI $400,000 to $910,000). From the healthcare sector perspective, the INMBs for an SSB tax, FV subsidies, and a combined policy were $1,933,000, $1,210,000, and $3,065,000, respectively, which suggested that thecombined policy was the most cost-effective compared to the status quo. From the societal perspective,the most cost-effective policy was the SSB tax.

Figure 3.2 shows the projected health and economic outcomes under each of the three policies in 10, 20, and 40 years as well as lifetime. As the number of years increased, the numbers of CHD and stroke events, CHD and stroke deaths, and QALYs gained increased steadily for each policy scenario. SSB taxes and FV subsidies could avert similar numbersof stroke events and deaths across different years, while SSB taxes could avert more CHD events and deaths compared to FV subsidies. The combined policy could avert more CHD and stroke events and their related deaths compared to implementing either of thetwo policies alone. However, SSB taxes could save more total costs compared to the other two policies.

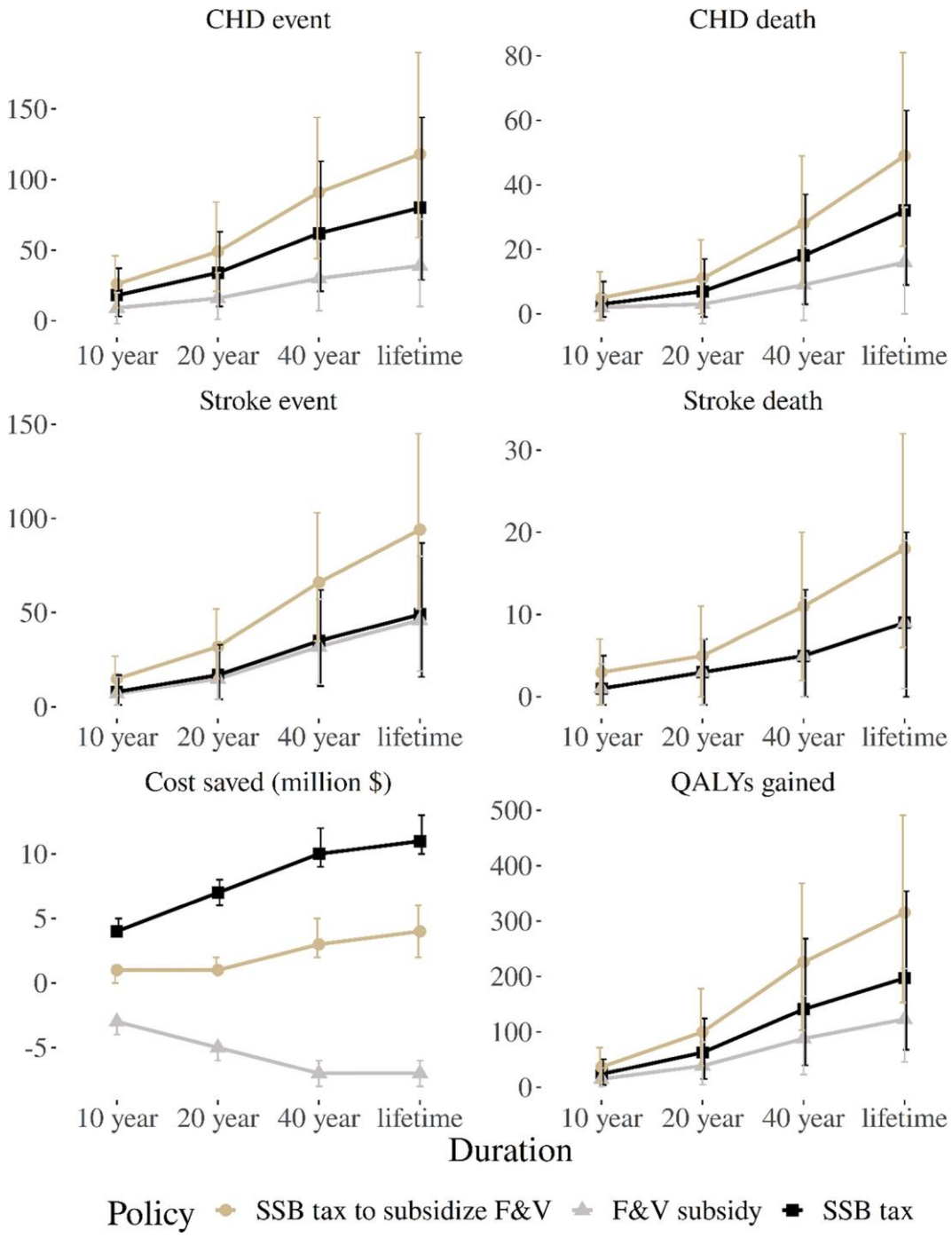


Figure 3.2. Projected long-term clinical and economic outcomes compared with the status quo (numbers of CHD and stroke events, QALYs, and healthcare costs per 10,000 adults in NYC over 10, 20, and 40 years and lifetime were reported).

Notes:SSB, sugar sweetened beverage; FV, fruit and vegetable; CHD,coronary heart disease; CVD, cardiovascular disease; QALY, quality-adjusted life year

Figure 3.3-4 shows the scatter plots of incremental costs and QALYs with 1,000 simulation iterations under each of the three policies from different perspectives. From a healthcare sector perspective(Fig. 3.3), the combined policy was dominant (i.e., reduced costs and increased QALYs) compared to all the other policies and had a 100% probability of beingthe preferred strategy at a $50,000/QALY cost-effectiveness threshold. However, from a societal perspective when policy costs were included, the combined policy was not cost-effective compared to the SSB taxes policy (Fig. 3.4). More specifically, the combined policy would result in an ICER of $268,462/ QALY compared to the SSB taxes policy. With a cost-effectiveness threshold of $50,000/QALY, SSB taxes alone had a 100% probability of being the preferred strategy from a societal perspective.

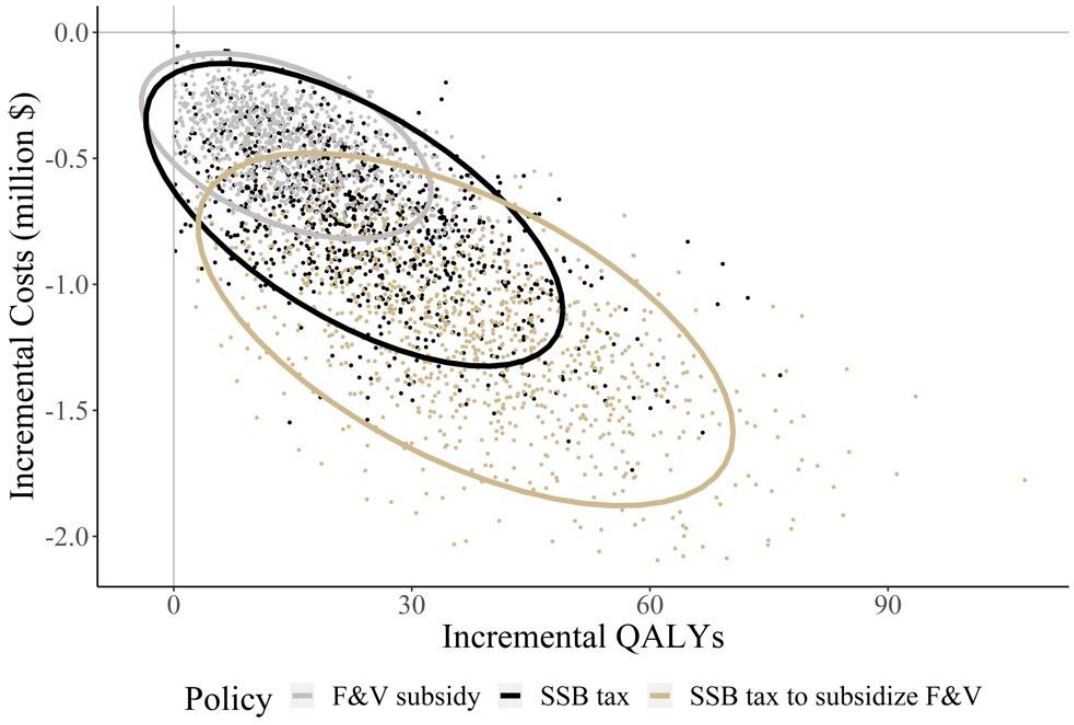


Figure 3.3 Incremental costs and quality-adjusted life years compared with the status quo: Healthcare sector perspective

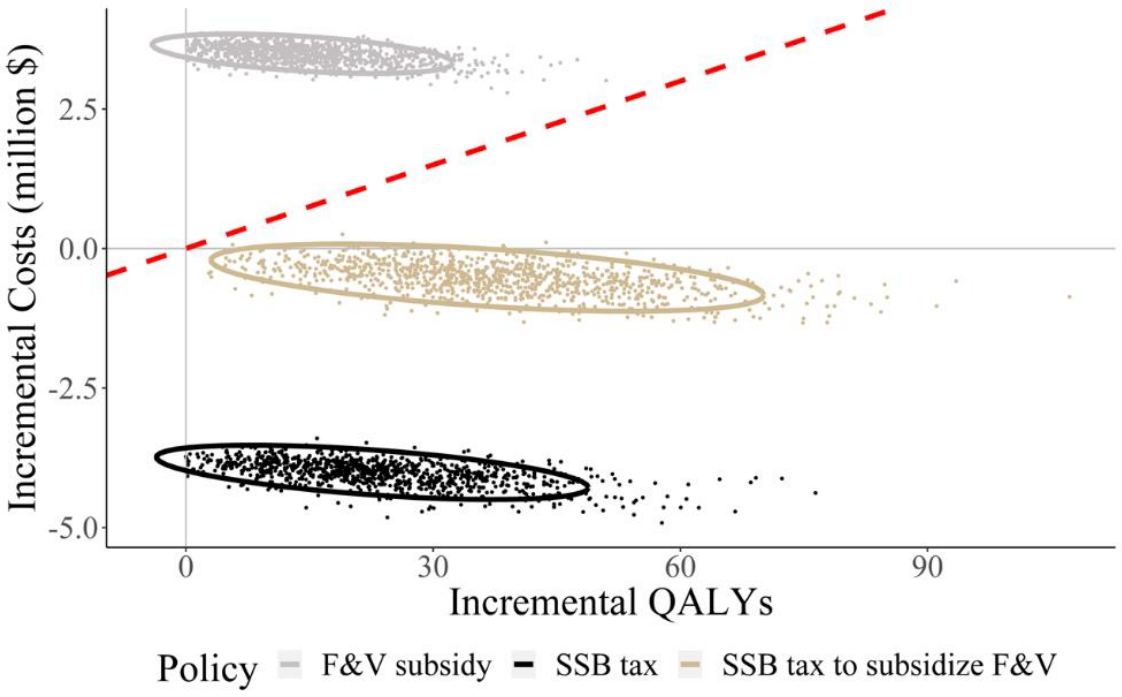


Figure 3.4. Incremental costs and quality-adjusted life years compared with the status quo: Societal perspective

Notes: The red dashed line is the $50k/ QALY cost-effectiveness threshold. SSB, sugar-sweetened beverage; FV, fruit and vegetable; QALY,quality-adjusted life year

Our subgroup analyses showed that the policy could have a differential impact across population groups by sex and race (Appendix B Table B.11). For example, funding FV subsidies with an SSB tax could avert 39 (95%UI, 9 to 73) CHD events and 18 (95%UI, 4 to 36) stroke events per 10,000 men compared to the status quo, while the averted cases for CHD and stroke were 16 (95%UI, 0to 36) and 13 (95%UI, 4 to 27), respectively, among women. The averted CVD cases by the other policies were also more pronounced among men compared to women. Our results also showed that Black adults were estimated to benefit more from the policies compared to White adults. For example, the combined policy could avert 34 (95%UI, 5 to 76) CHD events and 23 (95%UI, 0 to 48) stroke events per 10,000 Black adults, while the averted cases for CHD and stroke were 27 (95%UI, 0 to 57) and 14 (95%UI, 0 to31) per 10,000 White adults. In addition, the one-way sensitivity analysis on cost parameters shows that the cost-effectiveness of SSB taxes was more sensitive to the price and price elasticity changes compared with that of the FV subsidies policy (Appendix B Figure B.5).

3.4 Discussion

In this modeling study, SSB taxes, FV subsidies, or a combined policy where SSB taxes funded FV subsidies were projected to prevent a substantial number of incident CVD events and CVD deaths among adults in NYC. Our estimates equate roughly to 18,000, 11,000, and 29,000 CVD events being averted over 10 years in NYC if an SSB tax, FV subsidies, or the combined policy were implemented city-wide. The combined policy was estimated to also increase more QALYs compared to the other policies overtime. However, the cost-effectiveness of the combined policy depended on the specific perspective: the combined policy was the most cost-effective from the healthcare sector perspective, while from the societal perspective that considers both direct medical costs and policy and implementation costs, the most cost-effective policy was SSB taxes. In addition, the simulated policies could be more effective and cost-effective among men compared to women, and among Black adults compared to White adults. This implies that the policy has the potential to reduce racial and gender health disparities.

Several recent studies have assessed the cost-effectiveness of a potential national SSB tax policy or FV subsidy policy.[89,90] To our knowledge, our study is the first cost-effectiveness analysis of using SSB taxes to fund FV subsidies in a large city. Our study is particularly policy relevant because both quantitative and qualitative research indicates that SSB taxes garner greater support among stakeholders when emphasis is placed on using the revenue for health-related programs in the taxed communities.[91–93] Across the US cities that implemented SSB taxes, $135 million per year in revenue has been generated, with a substantial proportion spent directly on health-related programs.[59] Empirical studies suggested that SSB taxes discourage the purchase and consumption of SSBs.[94,95] Reducing SSB consumption through tax policy may have benefits beyond those captured in our analysis. For example, although evidence is still scarce, SSB consumption is associated with multimorbidity among adults,[96] and SSB taxes are expected to reduce the burden of such multimorbidity. A recent modeling study estimated that SSB taxes could reduce the burden of CVD and diabetes mellitus and result in substantial long-term healthcare expenditure savings in the USA.[90] Another modeling study projected that within one year an SSB tax policy could reduce mean BMI by 0.16 kg/m^2 among youth and 0.08 kg/m^2 among adults in the USA.[97] Furthermore, our analysis may be an underestimate of the long-term cost savings and health benefits of SSB tax policy because we did not include heart failure as an outcome, which

may be preceded by CHD and is associated with high costs and a significant reduction in quality of life. Notably, the ultimate impact of an SSB tax and FV subsidies is policy and context-specific; their effects may be dependent on factors such as the tax rate, baseline SSB and FV consumption, population demographics, and more.[59] This points toward the importance of evaluating the effects of a specific policy within its proposed location, using a model such as the one presented in this study.

NYC is a promising site for implementation of an SSB tax and FV subsidies. NYC recently received $5.5 million to expand the programs that offer FV financial incentives to SNAP recipients, [98] including Get the Good Stuff, which currently offers a one- to-one matching credit for eligible produce at six supermarket locations, and Health Bucks,[99] which offers coupons to reduce the cost of fresh FVs at all farmers markets. NYC also started offering a discount on prepackaged bags of locally grown produce purchased from local community-based organizations and urban farmers. As for SSB taxes, although more than 50 countries have implemented SSB taxes,[100] the USA has not implemented the policy nationwide because there has been opposition from the food and beverage industry against the policy,[101] and there are legal barriers such as state preemption.[102,103] Despite these challenges, several cities in the US and the Navajo Nation have implemented SSB taxes. NYC, however, lacks the authority to pass such a tax outright. The current study provides support for NYC to seek approval from the state legislature to grant the city the authority to pass such an excise tax and make decisions with respect to revenue allocation to ensure that at least a portion of the revenue is dedicated to low-resource communities in NYC to support FV affordability.

Funding FV subsidies with an SSB tax has the potential to reduce health disparities and address equity concerns associated with implementation of the tax. Low-resource communities and people of color experience higher rates of CVD and are disproportionately targeted through marketing by the SSB industry[104]. Concerns that an SSB tax would be regressive and weigh heavily on these communities are valid. However, young people and those with lower income have been shown to have a larger decrease in the purchase of SSBs as a result of a tax.[94,95] In this way, the potential health benefits from an SSB tax can be progressive.[59] To ensure that the greatest benefit goes to those currently bearing the greatest health burden from SSBs, the tax revenue should be dedicated to supporting the health of these low-resource communities and development and implementation

of such programs should be doneby guidance by community members themselves. In this way, programs will not only reflect the local needs of the communities themselves, but also build capacity and ideally, lead to longer-term sustainability and community-driven approaches.[57] Impactsof the tax and subsidy program on equity should be evaluated regularly, findings publicly reported and as needed, adjustments made, as possible. As such, an SSB tax and FV subsidies can serve as a strong strategy for reducing the burden of CVD and promoting health equity.

This study has several limitations. First, our analysis only focused on CVD-related health and economic outcomes, so the potential benefits of SSBtaxes and FV subsidies could have been underestimated. For example, research has shown that reduced SSB consumption and increased FV consumption are associated with a reduced risk of cancer,[105,106] which warrants further investigation into the effect of SSB taxes and FV subsidies in preventing cancer. Second, we assumed that CVD risk functions derived in national data would replicate the natural history of CVD among NYC adults. While it is difficult to test the validity of this assumption, our simulated mortality rates closely approximated those in NYC. Third, we did not model Latina/x/o or Asian Americans in the study due to poor representation and small sample sizes in the pooled cohort data and NYC HANES.Lastly, the nutrition policy environment is dynamic, and other local and federal legislation, as well as industry efforts, might be synergistic or antagonistic with SSB taxes or FV subsidies. However, we were not able to capture these potential complex interactions between the modeled policies and other nutrition policies in the current study.

Despite these limitations, to our knowledge, our study is the first assessment of the potential health and economic impact of combining an SSB tax with FV subsidies in a large city. The projected substantial health gains and cost saving associated with the policy could help relevant stakeholders and policymakers justify the implementation of this innovative policy in NYC and potentially other cities around the world.

4. ESSAY 3: 30-DAY HOSPITAL READMISSION PREDICTION AMONG OLDER ADULTS DISCHARGED TO SKILLED NURSING FACILITIES: AN INTERPRETABLE MACHINE LEARNING STUDY

4.1 Introduction

After an acute hospitalization for an illness or an injury, many older adults will require some kind of post-acute care, such as provided by a skilled nursing facility (SNF) to recover, improve functional status and manage chronic conditions. In the United States, approximately 20% of Medicare beneficiaries require SNF care after hospitalization, e.g., 1.7 million people received care from SNFs (2.4 million stays) in 2014 alone.[107] Spending on post-acute care, including SNF (which accounts for $28 billion annually) is over $60 billion per year in Medicare108 and is growing faster than inpatient spending.[109] A significant issue facing the post-acute care industry are rehospitalization rates or individuals returning to hospitals within 30 days of discharge—the costs associated with readmission burden the sustainability of the US healthcare system. In 2011, hospital costs related to readmission exceeded $41 billion.[110] Skilled nursing facilities (SNFs) have higher readmission rates than any other discharge location.[111] One in four patients discharged to an SNF is readmitted within 30 days.[112] Besides increasing costs, higher readmission rates are associated with higher patient mortality.[113] Higher readmission rates can also indicate poor care quality as the level of care in the post-acute facilities may lead to infections and complications.[114] Additionally, poor quality care can also harm the post-acute care market as patients can always find another place for care. SNF characteristics have been found to correlate with lower hospital readmission rates, suggesting the possibility that rates can be lowered by discharging to SNFs that specialize in post-acute care.[110,115]

Transitions between care settings have recently received much attention from researchers and policymakers.[116–118] For example, the Hospital Readmission Reduction Program is a CMS initiative that aims to reduce the number of rehospitalizations for older adults by tying rates to payment.[119,120] A significant component in the program's implementation is a model that establishes the expected number of readmissions within 30 days of discharge to assess which hospitals are having excess readmissions. From the hospital's perspective, there is a financial

incentive that is now coupled with the desire to provide patient-centered care. This increases the value of identifying which patients to be discharged are most at risk of readmission and what makes them at risk of readmission so that appropriate risk mitigating action can be taken.

One approach to improving rehospitalization rates is to focus on identifying what factors underly avoidable rehospitalizations. However, characterizing this subgroup from patient-level data has proven to be a difficult task.[121,122] In addition to patient-level characteristics another source of variability is the quality of post-acute care. Even among SNFs, care quality varies and several studies have found associations between facility-level characteristics and 30-day hospital readmission rates.[110,123] Further adding to the complexity is that preliminary evidence on social and economic factors of the patient has also been found to impact the risk of rehospitalization. However, such information is not always available from traditional data sources.[124] SNFs do play an important role in helping to stabilize patients discharged from hospitals thereby helping to avoid readmission, such that helping SNFs to target their residents who may need extra attention to stabilize their health could also be a viable strategy to reduce rehospitalization rates[125]. In this paper, we propose and compare several interpretable machine learning models built on data triangulated across patient, facility, and neighborhood socioeconomic factors to predict newly admitted SNF residents most likely to be readmitted to the hospital. Our objectives are to build both a highly predictive and interpretable model that will shed additional light on what factors most contribute to hospital readmission risk.

4.2 Related Literature

In this section, we review the studies that examine rehospitalization for SNF residents to highlight our contributions from three aspects: data used, methodology used to improve prediction performance and model interpretation.

The current literature tends to center on either insurance claims data (i.e., Medicare claims) or the MDS, both of which are collected at a national level by CMS. One of the more comprehensive studies was done by Neuman et al. who focused on Medicare beneficiaries receiving post-acute care from SNFs. They combined the resident-level data from the MDS, MedPAR, Medicare Beneficiary Summary files with facility-level characteristics using OSCAR and Nursing Home

Compare.[126] Chandra et al. focused on the SNF population but used resident-level data from the initial hospitalization (EHR and administrative datasets) for patients from a single hospital and 10 SNFs within an integrated healthcare delivery system.[127] Kimball et al. focused on Medicare enrollees who underwent total knee or hip arthroplasty, combining Medicare claims with SNF quality data from Nursing Home Compare.[128] Xu et al. combined data from the Minnesota Nursing Home Report Card (a state-level data source similar to Nursing Home Compare), the MDS, and Medicaid claims, to study the association between clinical care quality of SNFs and rehospitalization among Minnesota's 65+ Medicaid beneficiaries.[110] From a data perspective, our study stands out among those using the Minimum Data Set (MDS) and Nursing Home Compare to study the SNF population by augmenting this data with neighborhood-level socioeconomic and demographic data based on SNF location as well as including all resident regardless of payor source.

Concerning the modeling approach, most of the existing literature examines which resident and SNF characteristics are associated with hospital readmission using logistic regression, often focusing on certain sub-groups of residents.[114,123,127–130] Xu et al. took the approach a step further, developing a generalized linear mixed model, which combines the benefits of logistic regression for modeling binary outcomes with random effects used, in this case, to capture the correlation of residents within the same SNFs.[110] Neuman et al. used linear probability models, which have the simplicity of having coefficients on the original probability scale, but do not restrict the model's predicted probability to be between 0 and 1.[126] Kimball et al. examined the correlation between publicly reported indicators of SNF quality and time to 30- or 90-day hospital readmission using Cox-proportional hazard models.[128] Cox models have the advantage of informing if variables impact how long the event takes to occur. Several authors have attempted machine learning methods to take advantage of these methods' predictive performance at the cost of model interpretability. For example, Chandra et al. developed a gradient boosting machine (GBM) that achieved an AUC of 0.69 under 10-fold cross-validation, representing a 16% improvement over the Charlson Comorbidity Index method.[127] In this work, we focus on improving prediction by using machine learning and investigating the influence of imbalanced data, comparing variable selection techniques, and comprehensively comparing various machine learning models with improved prediction outcomes, and choosing the threshold for determining readmission cases.

Lastly, we apply methods to make the black-box machine-learning models interpretable to combine the benefits of accurate prediction with understandable risk profiles.

4.3 Material and methods

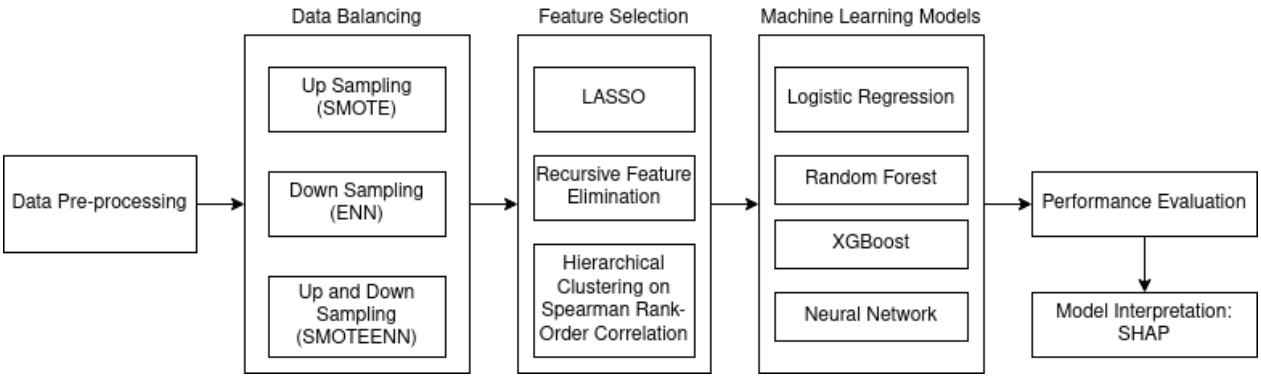


Figure 4.1. Outline of Analysis Approach

We first summarize our study framework (Figure 4.1). After data pre-processing, we apply three data balancing methods, a technique known to improve model predictiveness, resulting in three datasets. Then, we implement three feature selection methods on each of the three balanced datasets, resulting in nine datasets (3 balancing techniques by 3 feature selection techniques). Next, we train each of the four machine-learning models on the nine datasets. Finally, we evaluate the model performances and implement the model interpretation method on the best-performing classifier.

4.3.1 Dataset description

We combined three datasets in our research study: 1) Minimum Data Set (MDS) 3.0 (2014-2018) obtained from the Family and Social Services Administration of Indiana, US; 2) Publicly available SNF performance data (2015-2018) obtained from the Nursing Home Compare website published by the Centers for Medicare & Medicaid Services (CMS)[131]; 3) Socioeconomic Status and Demographic Characteristics of ZIP Code Tabulation Areas (ZCTA) data (2014-2017) obtained from publicly available National Neighborhood Data Archive (NaNDA).[132]

MDS 3.0 is a standardized core set of screening, clinical, and functional status elements of SNF residents. The data collection is federally mandated for all residents in SNFs that are Medicaid or Medicare certified. The data also includes administrative information, such as timing and location category of admission and discharge. Health and functioning information collected includes: cognitive patterns, e.g., the resident's attention, orientation, and ability to register and recall information; behavior, e.g., behavioral symptoms that may cause distress or are potentially harmful to the resident, or maybe distressing or disruptive to facility residents, staff members or the environment; functional status, e.g., activities of daily living (ADLs), altered gait and balance, and decreased range of motion; bladder and bowel, e.g., use of bowel and bladder appliances, the use of and response to urinary toileting programs, urinary and bowel continence, bowel training programs, and bowel patterns; active diagnoses, e.g., diseases that have a relationship to the resident's current functional, cognitive, mood or behavior status, medical treatments, nursing monitoring, or risk of death; health conditions, e.g., conditions that impact the resident's functional status and quality of life.

As MDS assessments are made of residents periodically, it is common for residents to have multiple records or data rows. To avoid double counting residents, we first extract each resident's admission record to serve as the denominator of the sample and to provide the resident features for prediction. Next, we check the records for each resident for the nearest discharge information (closest in time). If the nearest discharge assessment shows that a resident is discharged to an acute hospital and the length of the SNF stay is less than or equal to 30 days, then we denote this record as a readmission case. Otherwise, if the SNF LOS > 0, the discharge location is not to an acute hospital, or if there is no discharge assessment we denote this record as a non-readmission case.
The second data source, CMS's Nursing Home Compare website, is publicly available and contains measures of individual SNF performance. Performance indicators collected for this study are 18 clinical quality measures, 7 real-valued variables of expected and observed staffing intensity across four nursing roles, an ordinal summary rating of performance on state health inspections that ranged from one to five stars, and 9 binary variables indicating the SNF's ownership type. These data were merged into the MDS data at the SNF level.

The third data source, NaNDA, is also publicly available and contains measures of the physical and social environment linkable to other data by location information. From this data, we include 33 numerical features of the neighborhood's socioeconomic status and demographic characteristics. These numerical features are operationalized as proportions describing the population's income, education level, ethnicity, employment, and house ownership for a specific ZCTA. Data from the NaNDA were merged to SNFs in the MDS data at the zip 5 level.

The MDS includes data on all SNF residents for Medicaid and Medicare-certified SNFs. The inclusion criteria for our study are residents 65 years or older, admitted to SNFs from an acute care hospital, with the hospital discharge occurring between 2015 and 2018. This time range matched up well with our other two datasets at the time of the data pull. After filtering we have included 93,058 SNF residents across 368 Indiana SNFs. For features, we included 85 candidate features that should be available for prediction after the admission MDS assessment of the resident is completed.

4.3.2 Data imputation

We handle missing data in SNF resident features in two ways. First, we drop features with the proportions of missing greater or equal to 70 percent. Note that the label of a resident is always known. Then, we use the multivariate imputation method[133] to impute missing values for a given feature (Appendix C Table C.1 shows the missingness of the included features). This method works by building a regression model for each feature (y) with missing values as a function of other features (x), and uses the estimate y of for imputation.

4.3.3 Imbalanced data handling

The full dataset has a severe class imbalance or ratio of 7:1 in favor of the majority class. Class imbalance is known to degrade the performance of classifiers and various solutions have been proposed[134,135]. We tested several famous approaches that treat the imbalance problem from different angles to assess the impact of the balancing technique on the final performance of our classifier models. We used SMOTE,[136] an oversampling technique, to create new synthetic samples based on the minority class. For this algorithm, the majority class is unaffected and the

synthetic new occurrences are different from minority cases that already exist (i.e. they are not simple duplicates of observed cases). The second algorithm, ENN is an under-sampling technique that clusters data based on the feature space and then deletes the majority-class cases with a different observed output value than their nearest neighbors.[134,137] The third balancing algorithm, SMOTEENN,[134] is a combination of both of the other two. The algorithm uses SMOTE to oversample minority cases and ENN to remove majority class samples from the dataset. We use Python "imblearn" package to implement all data balancing techniques, and we have a 1:1 ratio for the two classes after data balancing.

4.3.4 Feature selection

As the feature space was large, we implemented three feature selection methods to reduce overfitting, and compared them with using all features. The feature selection methods we adopted are 1) LASSO logistic regression,[138] 2) recursive feature elimination (RFE) using random forest with cross-validation, 3) hierarchical clustering of highly correlated features using Spearman rank-order correlation. The LASSO is a penalized regression method that biases some coefficients among correlated features to zero, eliminating the corresponding features from the model and is analogous to specifying a Laplace prior for the coefficients in a Bayesian regression model.[139] The RFE with cross-validation searches for a subset of features by starting with all features in the training dataset and removing features until the desired number remains. We adopted hierarchical clustering of highly correlated features since the first two feature selection methods eliminated a limited number of features (LASSO 1 and RFE 13). Several features in the data are collinear (e.g., various measures of facility size such as number of admissions, number of staff, and number of beds), such that permuting one feature will have little effect on the model's performance as the model can get the same information from a correlated feature. Note that we do not present results from LASSO feature selection since the number of features after selection is almost identical to the number of all features, making the results trivially different from the benchmark case.

4.3.5 Machine learning classifiers and evaluation metric

We utilized four machine learning classifiers to predict hospital readmission: logistic regression, random forest, XGBoost, and neural networks. All models were fit with the Python Sklearn or xgboost (in the case of XGBoost) package using default settings.

Logistic regression (LR)[140] is a commonly used model for binary classification tasks. It uses the sigmoid function to learn the linear relationships between the outcome variable and the independent variables. As a type of generalized linear model, features used in logistic regression have interpretable parameter estimates (e.g., odds ratios) that also permit statistical hypothesis testing. Additionally, LR permits the estimation of interaction effects between features (sometimes referred to as moderating effects). We use the performance of LR as a benchmark for other approaches.

Random forest (RF)[141] builds decision trees out of a subset of samples and features from the original training dataset and classifies them based on the majority vote from each tree. The use of multiple trees helps to reduce classification errors and guard against the impact of noise in the data. RF also has internal measures of variable importance.

XGBoost, or extreme gradient boosting of trees,[142] iteratively trains an ensemble of shallow decision trees parallelly. In each iteration, it uses the error residuals from the previous model to fit the next model. The final prediction is a weighted sum of all tree predictions. XGBoost has similar measures of feature importance as RF. Compared to RF, XGBoost is more computationally expensive and requires more parameters to tune.

Neural networks[143] are designed to mimic the way the human brain processes information. They are comprised of input (independent variables) and output (dependent variable) nodes with connections between them called edges. Edges are learned by using connection weights, bias weights, and cross-entropy. Additionally, one or more hidden layers can be placed between the input and output layers in the neural network framework.

Table 4.1. Confusion matrix

		Actual Values	
		Positive	Negative
Predicted Values	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

For each model, 70% of the data is randomly chosen as the training data while 30% of the data is held out as test data. Model performance is evaluated based on several metrics calculated from the model's predictions on the test data. We report on many of the classic binary prediction metrics such as accuracy, precision, recall (sensitivity), specificity, F1-score, ROC curve AUC, and precision-recall curve which are all derived from the confusion matrix described in Table 4.1. More specifically, TP (true positive) is the number of cases that the model correctly predicts the positive class. TN (true negative) is the number of cases that the model correctly predicted the negative class. FP (false positive) is the number of cases that the model falsely predicts the positive class, and FN (false negative) is the number of cases that the model incorrectly predicts the negative class.

Accuracy is the proportion of correct predictions. However, it is susceptible to overstating the benefits of a prediction model when the two classes are imbalanced.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \qquad \text{Equation 4.1}$$

In readmission prediction, false positives will lead to unnecessary and excessive use of healthcare resources that are often scarce. Thus, making correct predictions for positive cases is more important, and precision fits this purpose by calculating the proportion of correctly forecast positive observations over all cases predicted positively (i.e., when the model predicted positive, how often was it correct).

$$Precision = \frac{TP}{TP+FP} \qquad \text{Equation 4.2}$$

Recall or sensitivity calculates the proportion of true positives across all actual true positives (i.e., when the model should have predicted a positive, how often did it predict a positive).

$$Recall = \frac{TP}{TP+FN} \quad \text{Equation 4.3}$$

Specificity is a complementary measure of recall or sensitivity. It is the percentage of predicted negatives across all actual negatives (i.e., when the model should have predicted a negative, how often did it predict a negative).

$$Specificity = \frac{TN}{FP+TN} \quad \text{Equation 4.4}$$

An effective readmission prediction model should both identify positive cases when they occur and when the model predicts a positive case, it should be truly positive a high percentage of the time. Precision captures the first idea and recall or sensitivity captures the second. The F1-score accounts for both as it is the harmonic mean of the two measures and is calibrated to work well on imbalanced data. It has a maximum value of 1, meaning perfect precision and recall, and a minimum value of 0 when either precision or recall is zero.

$$F1 - score = \frac{2(Recall \times Precision)}{Recall+Precision} \quad \text{Equation 4.5}$$

Lastly, the receiver operator curve (ROC) area under the curve (AUC) measures the tradeoff between sensitivity and specificity across the range of possible probability cut-off values for predicting a positive or a negative case. AUC ranges from 0.5 to 1, with higher being better.

4.3.6 Model interpretation method

One of the drawbacks of more flexible machine learning approaches to prediction is the difficulty in interpreting why the model performs well or what variables in particular are contributing to correct predictions. SHAP (SHapley Additive exPlanations)[144] is a method to explain individual predictions that takes a game theory approach to divide the contributions among the features for

an individual predicted value. The approach uses Shapley values,[145] which are derived from contributions of players cooperating in a game with competition between groups of players.

The calculation of Shapley value is performed for feature *X* of some instance i. With a given feature subset, the SHAP method will compute the differences between the predicted outcomes of instance i including and excluding the feature X. Such difference is the marginal contribution of the given feature X to the current feature subset. Then this calculation is performed for all feature subsets (all possible feature combinations) including and excluding the feature X. The weighted mean of the differences is the Shapley value for instance i. If we have a data set with *N* features, then the SHAP method needs to train 2^N models with the same parameter settings. Below we show the mathematical formulation for the Shapley value:

$$Shapley_X(i) = \sum_{s=\{s \vee X \in s\}} \frac{P_s(i) - P_{s-X}(i)}{|s| \times \binom{F}{s}} \qquad \text{Equation 4.6}$$

where *s* is a subset of the features including feature *X*, $P_s(i)$ is the predicted outcome from a model trained with the feature subset *s* for instance i, $|s|$ denotes the size of a feature subset, and *F* is the total number of features.

For our work, we applied the SHAP model to first generate SHAP values for all cases in our dataset and then illustrated the SHAP summary plot and SHAP dependence plot using a publicly available SHAP API [40] to gain a global understanding of our dataset. We implemented the SHAP method on the best-performing classifier (XGBoost), and the SHAP value is generated for each instance.

4.4 Results

4.4.1 Overall Performance

In this section, we compare performance results across classifiers, sampling schemes, and feature selection methods. Note that we omitted the performances of RF classifiers since their performances were similar to XGBoost but being an average of 4% inferior to the XGBoost classifier across all performance metrics. Also, we did not present the results using LASSO feature selection which removed only 1 feature making results redundant to the baseline case.

Figure 4.2 compares the classifiers across the feature subsets when sample balancing was fixed to be SMOTEENN (the best performing of the sample balancing techniques). The figure shows that the XGBoost classifiers perform the best with AUC of 0.99 across all feature settings whereas the other two classifiers have a much greater decline in AUC when reducing the number of features in the model. LR classifiers perform the poorest among the candidate methods when using the same features. LR classifier was also the most affected by the reduced size of features as AUC goes from 0.92 to 0.79.

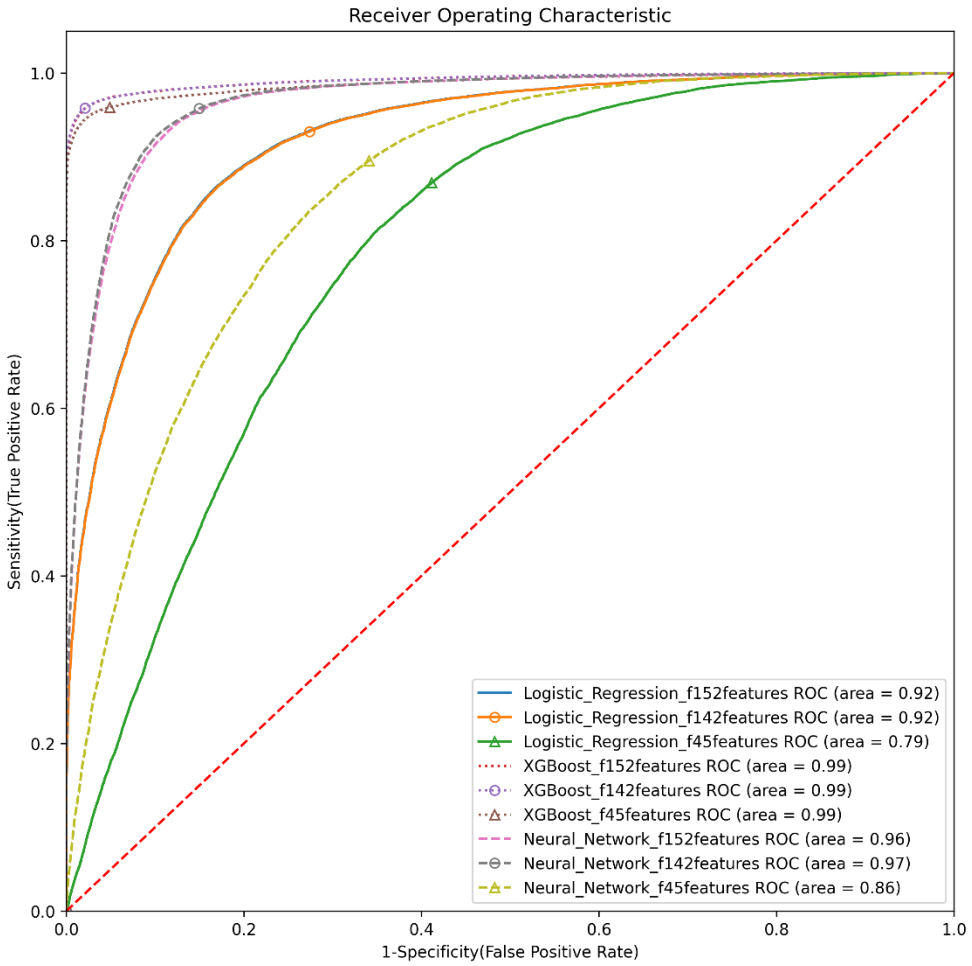


Figure 4.2. ROC Curve Comparing Classifiers' Performance with SMOTEENN (up-and-down sampling)

Table 4.2 contains the overall performance measures of precision, recall, F1-score, and accuracy for the three classifiers with SMOTEEN balanced data and the prediction threshold value is set to

0.5, i.e., a prediction is labeled as readmission if the predicted probability is larger than or equal to 0.5 and vice versa. The precision, recall, F1-score and accuracy for LR classifiers decrease when the number of features decreases, while XGBoost and NN both have a slight increase in these metrics when 10 features were dropped using RFE. When using the correlation and clustering feature selection method, the recall, F1-score and accuracy for the XGBoost classifiers only drop by about 1%, which is more robust compared to the NN classifiers (performance metrics drop more than 10%). XGBoost performs the best with the smallest number of features (45) with precision of 98%, recall of 93.9%, F1-score of 95.9%, and accuracy of 96.4%.

Table 4.2. Performance Metrics for Classifiers with Threshold

Model	# of features	Precision	Recall	F1-score	Accuracy
SMOTEENN (up and down sampling)					
Logistic Regression	152	81.5%	85.8%	83.6%	84.6%
	142	81.4%	85.8%	83.5%	84.6%
	45	66.8%	78.2%	72.1%	72.3%
XGBoost	152	97.9%	95.3%	96.6%	96.9%
	142	98.0%	95.4%	96.7%	97.0%
	45	98.0%	93.9%	95.9%	96.4%
Neural Network	152	87.7%	92.4%	90.0%	90.6%
	142	88.1%	92.8%	90.4%	91.0%
	45	73.2%	80.2%	76.5%	77.6%

Note: The 152 features represent the full feature set, the 142 features are chosen by RFE and the 45 are selected by hierarchical clustering of highly correlated features. RF classifier model is omitted as its performance closely mirrored XGBoost but was about 4% worse across metrics.

Next, we compare the XGBoost classifiers' performances across the three data balancing/sampling schemes. Note that we choose the XGBoost classifiers with hierarchical clustering of highly correlated features to simplify the presentation of results as the pattern holds for the other classifiers and feature selection methods. Figure 4.3 shows that the XGBoost classifier performed the best with an AUC of 0.99 using the up-and-down sampling scheme (SMOTEENN). The upsampling scheme (SMOTE) has 0.04 lower AUC, and the downsampling scheme (ENN) has 0.14 lower AUC than using SMOTEENN. Table 4.3 compares data balancing approaches using precision, recall, F1-Score, and accuracy. Across all metrics, the up-and-down sampling performs

the best (i.e., about 5% better on average than upsampling and about 20% better on average than downsampling). Thus, we would recommend using SMOTEENN since it has the best performance across all metrics, especially when we concern more about recall, and it is more efficient than SMOTE since it will need to generate fewer samples, and synthesizing new samples take more time.

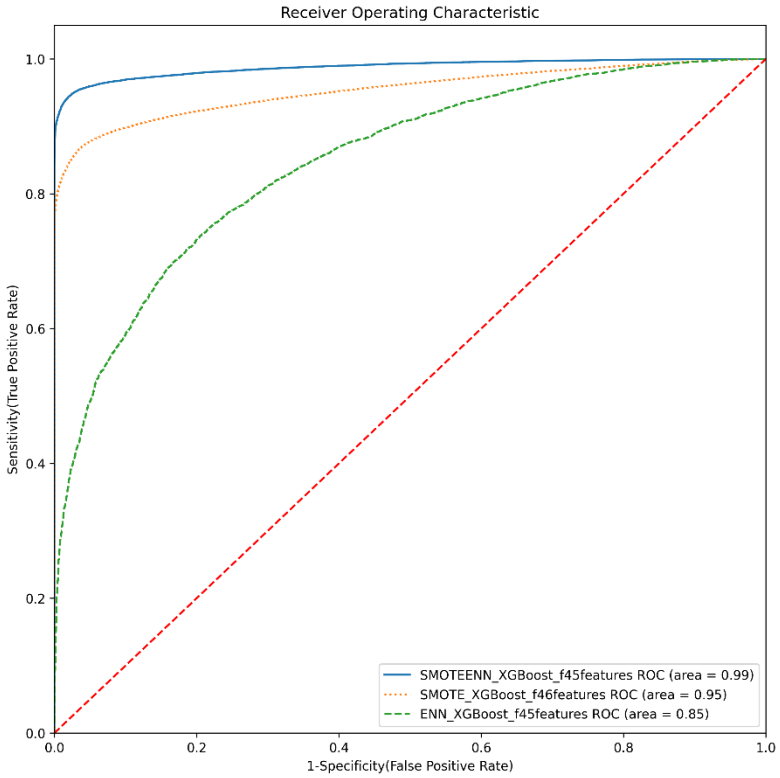


Figure 4.3. ROC Curves Comparing Classifiers' Performance under Three Sampling Schemes

Table 4.3. Performance Metrics for Classifiers with Three Data Balancing Methods (Threshold of 0.5) for Readmission Cases

Model	# of features	Data balancing	Precision	Recall	F1-score	Accuracy
XGBoost	45	SMOTEENN (up and down sampling)	98.0%	93.9%	95.9%	96.4%
	46	SMOTE (up sampling)	96.0%	86.7%	91.1%	91.5%
	45	ENN (down sampling)	78.8%	72.5%	75.5%	76.5%

Note: Pattern seen in this table holds across other classifiers and feature selection methods.

Table 4.4 considers changes in the performance metric due to varying the choice of the threshold value for prediction. We focus on the XGBoost classifier with 45 features since they perform better than other classifiers using fewer features across all metrics. In Table 4.4, when the threshold value is set to 0.5, the XGBoost classifier using 45 features has precision of 98%, recall of 93.9%, F1-score of 95.9%, and accuracy of 96.4%. When the threshold value is set to 0.44 to maximize the F1-score, we found improvement in F1-score and accuracy to be slight, while recall increases by 0.4% and precision falls by 0.4%. As we cared more about recall since not detecting readmission cases will be more costly in healthcare resources, and even if we do not have a large increase in recall, the impact would be substantial when the entire population becomes large. Note that we also tested to maximize recall, but that approach resulted in a threshold close to 0, labeling all cases as readmitted while making wrong predictions for all non-readmitted cases. Thus, the superior approach was setting the threshold value to maximize the F1-score, which resulted in a higher recall for the classifiers with good TN.

Table 4.4. Performance Metrics for Classifiers with Varying Thresholds Readmission Cases

Model	# of features	Threshold value	Precision	Recall	F1-score	Accuracy
SMOTEENN (up and down sampling)						
XGBoost	45	0.5	98.0%	93.9%	95.9%	96.4%
		0.44	97.6%	94.3%	95.9%	96.4%

Note: Threshold value of 0.44 was chosen to maximize the F1-score.

4.4.2 Model Interpretation

We use the SHAP method to interpret the results from our best-performing model, the XGBoost classifier, and focus on the version trained on the least number of features (45). Figure 4.4 is the SHAP value summary plot, with each row displaying the distribution of each feature's SHAP values. The color of a data point for an instance indicates the value (i.e., for continuous features higher values are redder, and lower values are bluer). For example, higher values of ADL tend to either have a large positive impact on the predicted probability (the red values on the far right of the ADL row) or small to moderate negative values. Looking at features all at once allows for a relatively quick diagnosis of which features are positively or negatively correlated with predicted readmission probability and which features may have more complex relationships, such as an interaction effect.

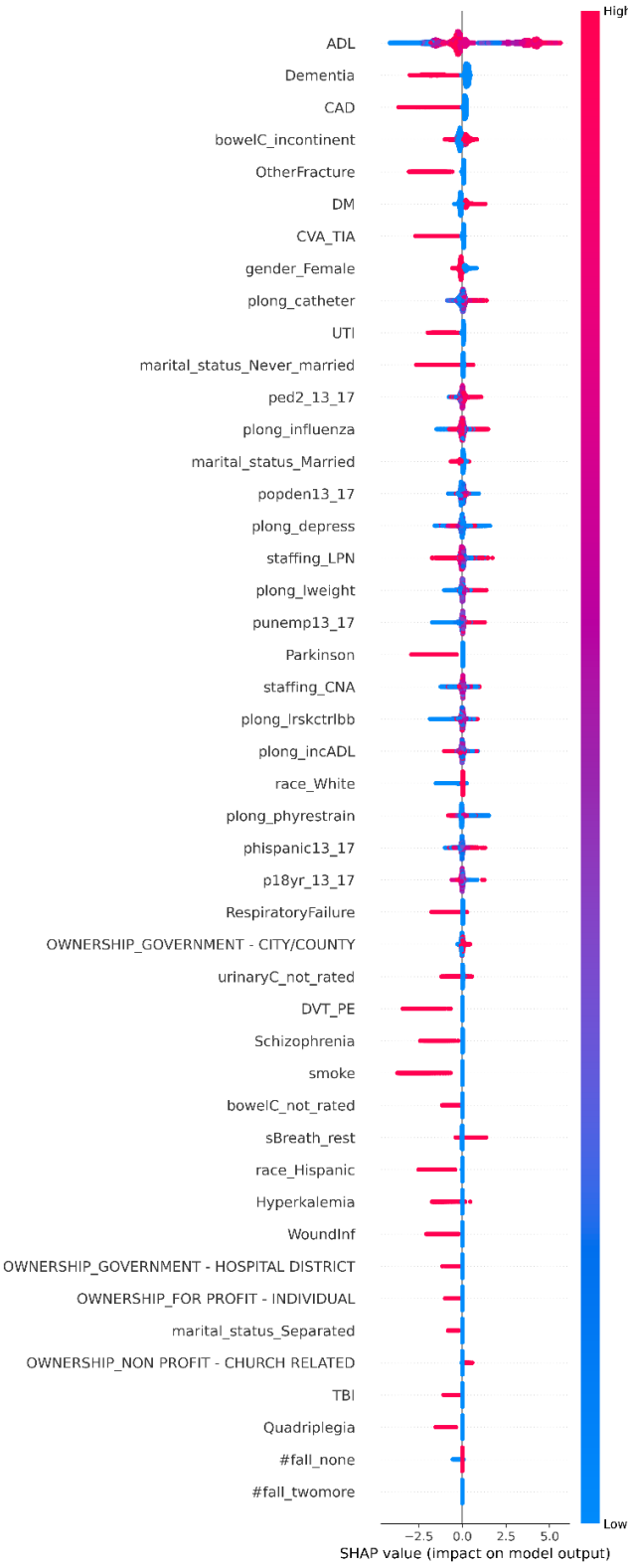


Figure 4.4. SHAP summary plot for XGBoost model with 45 features

We report and discuss our findings on what features contribute to lower readmission risk based on these SHAP values (Figure 4.4). For example, the resident's average SHAP value was negative for individuals diagnosed with dementia (feature value of 1), and positive when the resident was not diagnosed with Dementia (feature value of 0). That is, in the SHAP interpretation for our prediction model, readmitted patients tend not to have dementia. We organize the following results from the SHAP values around three categories: resident, facility, and neighborhood features. Several resident-level features were found to be positively correlated with the prediction of readmission. These resident features include Diabetes Mellitus diagnosis, shortness of breath or trouble breathing when sitting at rest, and being male. Resident features that contributed to the lower prediction of readmission probability were diagnoses for coronary artery disease (CAD), fractures other than hip fracture, cerebrovascular accident (CVA), transient Ischemic attack (TIA) or stroke, urinary tract infection, Parkinson's disease, deep venous thrombosis (DVT), pulmonary embolus (PE) or pulmonary Thrombo-Embolism (PTE), wound infection, traumatic brain injury, and quadriplegia. Tobacco use, being married and married but separated, and Hispanic ethnicity were also correlated with lower predicted readmission probabilities. The relationship between ADL need for assistance and predicted readmission probability follows a U-shape. Both High ADL scores and low ADL scores are related to higher predicted probabilities of readmission risk. This may be due to the correlation between ADL and other features.

Facility features that were positively correlated with predicted readmission probability include a higher percentage of long-stay residents with a catheter inserted and left in their bladder, a facility owned by city/county government (small impact), and a facility owned by church-related entities (non-profit). Facility features correlated with lower predicted probabilities of readmission were having a higher proportion of residents under the age of 18 (small impact), having a higher percentage of long-stay residents who were physically restrained, having a higher percentage of long-stay residents who lose too much weight, and ownership structure being by the government (hospital district) or by a for-profit organization.

Neighborhood features that were positively correlated with predicted readmission probability were a higher proportion of individuals with high school diplomas and/or some college, a higher proportion of individuals age 16 or older who were in the labor force but unemployed, and a higher proportion of Hispanic residents.

Figures 4.5-8 are used as examples to examine features for non-linear relationships with predicted readmission probability and interactions between features by using dependence plots. For each plot, the x-axis shows the feature value, the left y-axis shows the Shapley value of the same feature, and the color bar on the right side of the plot shows a second feature's value. The dispersion of the data points occurs when multiple instances have the same feature value (but not Shapley value) for the feature on the x-axis. We can understand such change as the deviation of predicted readmission risk from the mean of prediction risks using all possible combinations of features. The further a point deviates from the mean of predictions (which is 0), the more impact the features have on the prediction in that instance. Therefore, a positive correlation between the feature on the x-axis and the predicted readmission probability would be seen as a tendency to have positive SHAP values on the right side of 0 on the *x*-axis and negative SHAP values on the left side of the mean of the x-axis. Similarly, if red data points tend to have positive SHAP values and blue data points tend to have negative, then the feature on the right y-axis is positively correlated with the prediction of readmission. If the relationship between data point color and the SHAP values changes as the x-axis variable increases, this would indicate that there is an interaction effect between the two features and the predicted probability of readmission. Figure 4.5 shows that an SNF located in a neighborhood with a low unemployment rate (less than or equal to 0.2 after normalization) is associated with lower predicted readmission probabilities. As the neighborhood's unemployment rate increases, the predicted probability of readmission increases in a non-linear pattern. The pattern is such that the increase of the predicted readmission probability grows quickly at first when the proportion of unemployment is low but rising, runs flat over middle values of unemployment, and grows slowly when the proportion of unemployment becomes high. Using the color coding of data points for gender, we find that for neighborhoods with a low proportion of unemployment, female residents are associated with even lower predicted readmission probabilities than males (an interaction effect). This pattern is similar to that seen in Figure 4.6 for the feature low percentage of low-risk long-stay residents who lose control of their

bowels or bladder. In addition, individuals from a minority race or ethnicity at facilities with a high value on this feature are predicted to have relatively higher predicted readmission risk. Figure 4.7 shows a polynomial relationship between satisfied staffing of Certified Nursing Assistants (CNA) and predicted readmission risk. Low and high values of satisfied staffing of CNA indicate either unmet or excessive CNA staffing hours respectively, are associated with lower predicted readmission risk. It is possible that CNAs are mainly responsible for more basic needs care, which is less important to post-acute patients who need more medical care.

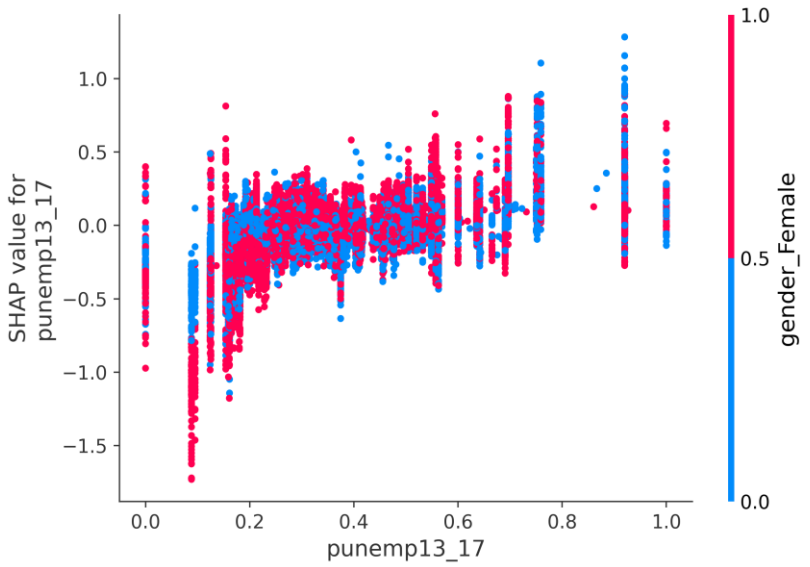


Figure 4.5. Impact on readmission risk of SNF neighborhood unemployment rate and resident gender

Note: Red points indicate that the resident is female and blue that the resident if male.

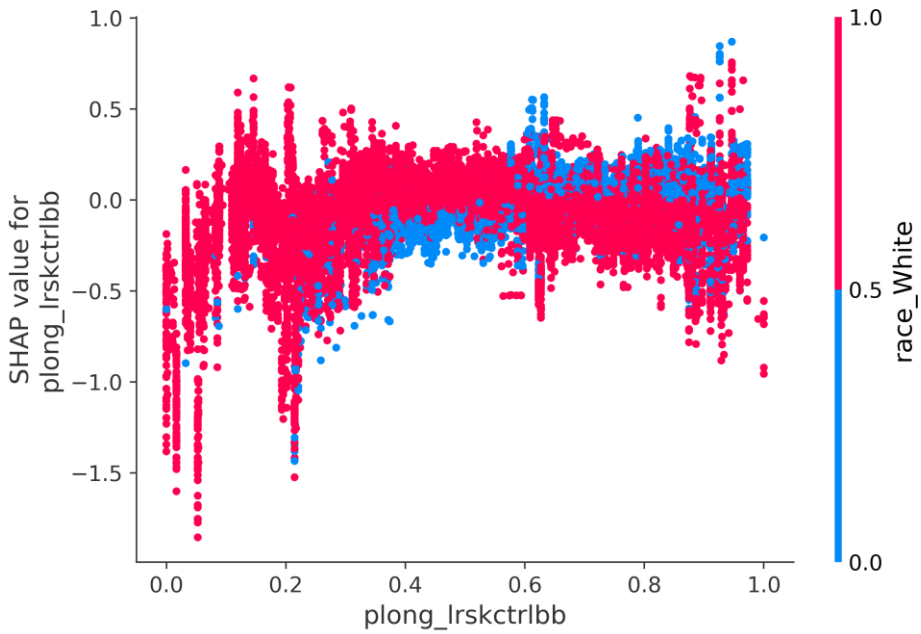


Figure 4.6. Impact on readmission risk of percentage of low-risk long-stay residents who lose control of their bowels or bladder

Note: Red points indicate that the resident identified White as their race and blue points that the resident identified any other group as their race.

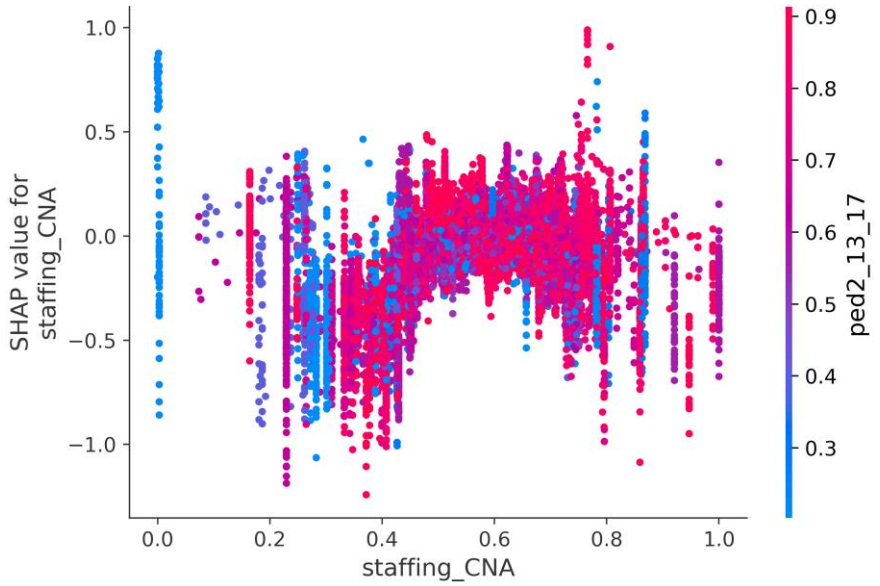


Figure 4.7. Impact on readmission risk of staffing of Certified Nursing Assistants (CNA)

Note: The color of the points indicates the normalized proportion of population received high school education

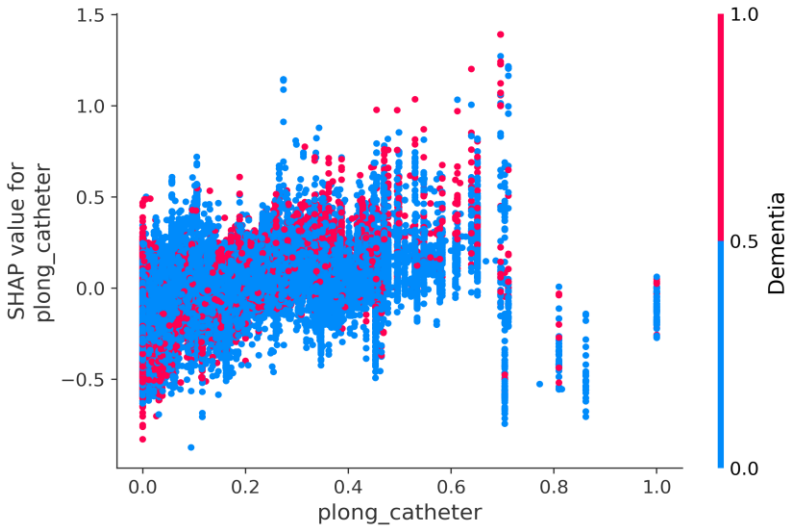


Figure 4.8. Impact on readmission risk of percentage of long-stay residents with a catheter inserted and left in their bladder

Note: Cases are stratified by dementia diagnosis such that red indicates a diagnosis of dementia and blue indicates no diagnosis.

4.5 Conclusion

In this paper, we propose and compare several interpretable machine learning models built on data triangulated across patient, facility, and neighborhood socioeconomic factors to predict newly admitted SNF residents most likely to be readmitted to the hospital. We find that careful handling of data imbalance and feature selection improves machine learning model performance. More specifically, the XGBoost classifier using up-and-down sampling and feature selection with hierarchical clustering outperforms other models with AUC of 0.99. With the SHAP model interpretation method, we found that the neighborhood social-economic factors correlate with the readmission risk, such as the neighborhood's unemployment rate around the SNF. One application of this tool could be developing a recommendation list of nursing facilities for hospital discharge planners. It could serve as a risk-adjustment model in evaluating nursing facility success in minimizing re-admissions after considering the risk profile in their patient populations.

A limitation of this study is that our data does not comprehensively capture the patient's health history before their admission to SNFs, especially during their most recent hospital stay. Existing literature shows that such variables could have statistical significance. We also normalized real-valued features to range between 0 and 1, which makes interpreting results less convenient, but this could be addressed by reversing the normalized values. The data for the study comes from only one state which may limit the generalizability of the model or results. However, the data from this study exist at a national level and future work should focus on testing the current model on data occurring later within the same state and on data from other states.

The developments in interpretable machine learning illustrated in this study highlight the approach to improving trust in machine learning-based prediction models in clinical practice. We significantly improved the predictive performance of the readmission risk model over the logistic regression model without losing all of the interpretability that has long been the advantage of parametric statistical methods. For future research, developing automated methods to identify meaningful relationships between risk factors and the readmission outcome that do not require manually building and reviewing plots should profoundly impact the uptake of interpretable machine learning models. From the problem context perspective, we will also investigate the interpretable machine learning tool to address the multiple interdependent outcomes for the SNF residents, including death, general hospital readmission, and discharge back to the community, all of which are relevant to the SNF.

5. CONCLUSION

Predictive modeling shows great promise in assisting with decision-making problems in healthcare, particularly with the rise of big data and computational technology. As more healthcare domain researchers and practitioners understand the power of predictive modeling and as healthcare data pipelines improve, developing predictive modeling tools to support decision-making in healthcare is a growing area of research. To achieve optimal results, we still need more tailored innovative methods to process and analyze complex data in health analytics applications.

My dissertation focused on using predictive modeling techniques to address specific challenges and assist with complex decision-making problems in three healthcare domains: clinical practice, public health, and health systems.

Paper 1 focuses on the domain of clinical practice. It introduces a novel predictive modeling method tailored to medical practitioners' needs to identify the most cost-effective liver cirrhosis diagnostic strategy with limited data access. This model also captures the heterogeneity of the various diagnostic strategies' outcomes. The advantage of such a predictive model is that it aims to support practitioners' decision to select the most cost-effective diagnostic strategy in a timely and low-cost fashion. The development of such a model involves expert clinical knowledge. This model does not require individual patient data which avoids any delays related to data acquisition due to data privacy regulations. The model allows practitioners to identify a few cost-effective strategies quickly. This helps the practitioners narrow down their choices of strategies, and they can further design clinical trials for research on targeted strategies with less time and lower costs.

Paper 2 presents an individual-based CVD simulation model to evaluate public health outcomes and healthcare costs using various food policies. The modeling of individual disease progression is transparent, easy to understand, and verified by cardiologists. The model captures individual heterogeneity, such that the probability of having CVD depends on multiple patient characteristics. Meanwhile, it can also quickly output various outcomes for a relatively large population. With the assistance of this predictive model, policymakers can get reasonably quick feedback on the likely consequences of different potential policies. This quantitative evidence will help them make a

more informed decision about which policy can improve public health with reasonable costs from a pool of candidate policies implemented within a complex population. Moreover, the CVD predictive model can also offer outcomes according to subpopulation groups (e.g., by income, race, etc.). This is particularly important if health policy aims to achieve health equity, or at least reduce health disparities, since policies implemented within a region usually influence multiple population subgroups differently (e.g., across socio-economic levels or race and ethnicity groups).

Lastly, paper 3 took a different perspective within a specific healthcare system between a hospital and SNFs. Both hospitals and SNFs will suffer from high readmission rates because the readmission rate is an essential metric for quality of care. Low quality of care can make SNFs less competitive in attracting new admissions, particularly in markets where post-acute care supply outpaces demand. Additionally, the rise of value-based payments, or payment rates tied to quality metrics, increases the incentive of SNFs to reduce hospital readmissions to protect revenue. For example, hospitals and SNFs are penalized for high readmission by the Centers for Medicare & Medicaid Services with respect to Medicare payment rates. Thus, there is some incentive for hospitals and SNFs to coordinate to minimize patients' readmission risks. As an initial step, paper 3 introduces a readmission prediction tool using machine learning methods by augmenting multiple data sources, including patient demographic, health, and functioning data from the MDS, SNF characteristics from Nursing Home Compare and SNF neighborhood characteristics data. The advantage of the machine learning techniques is that they significantly improved prediction accuracy by relaxing the linear assumption of more traditional parametric statistical methods. Additionally, I illustrated how the ubiquitous black-box machine learning models could be better understood with a model interpretation method. This prediction tool can further contribute to the study of decision-making on patient discharge assignment recommendations within the hospital-SNFs system.

This dissertation explores predictive modeling in three healthcare domains: clinical practice, public health, and at the health system level. However, there are many research opportunities to improve predictive modeling to better address the needs of various stakeholders in an era in which technology in healthcare rapidly changes. I list some opportunities from the three healthcare domains mentioned. From the clinical practice perspective, many wearable health devices and

sensors have been deployed and such data accumulate in real-time. How to develop predictive models to augment such information with existing clinical practice to facilitate medical practitioners' decision-making in personalized medical treatment is an exciting topic. From the public health perspective, predictive modeling needs to be updated promptly for properly evaluating new policies as populations change due to many factors, such as changing health behavior, diet, work-life style, and climate change. Thus, predictive models developed previously may not serve a later population at a later time. Manually re-calibrating and validating predictive models can be tedious and time-consuming. Methods to automate predictive models' calibration and validation processes are worth exploring. Lastly, in the healthcare systems domain, and particularly the system with both hospitals and SNFs, we can further answer the question "Which SNF should a hospital discharge case manager recommend for a post-acute patient who needs SNF care". We can tackle the problem by developing a simulation model to study the coordination between a hospital and multiple SNFs on patient discharge assignment recommendations.

In addition to the potential research opportunities, I will reflect on my journey in developing predictive modeling in healthcare based on my experience collaborating with multiple medical researchers in many prestigious medical institutions. The most important factor is communication. We (researchers developing predictive modeling tools) must communicate regularly with the stakeholders to understand the healthcare problem context and their needs. Often, there are bi-directional knowledge gaps between the stakeholders and us. For example, we are more knowledgeable about predictive modeling techniques but less familiar with the healthcare problem context and the intricate details of the healthcare problems. Meanwhile, the stakeholders usually do not follow the most advanced predictive modeling techniques. Still, they need to understand the methods for solving their problems since the stakeholders (e.g., physicians) hold liabilities to their clients. This communication process often takes time and goes back and forth due to a better understanding of the problem context and the modeling techniques. Finally, we may agree on a different method for a better-framed problem. The key factors in communication include always clarifying the needs of the stakeholders (e.g., identifying the essential metrics to capture and what needs to be compared), knowing how the data was collected, conceptually explaining why using a particular model, and using intuitive visual presentations for the results. Do not go into the depths of technical details until there is any concern from the stakeholders.

One major challenge in developing predictive modeling for healthcare problems is data access. Healthcare problems are often complex, and a predictive outcome could be correlated with many factors in a healthcare problem. One major factor is patient characteristics data; sometimes, we need data from multiple care settings (e.g., different hospitals, skilled nursing facilities, etc.) and insurance companies. Due to privacy and law restrictions, we often cannot access such high-resolution data to develop a very sophisticated model that closely captures a real-world system. However, we can still develop useful predictive models to meet the stakeholders' needs. Predictive models offer a wide range of methods, and some do not require high-resolution individual patient-level data. Instead, we can acquire low-resolution population-level data from the literature. This approach is suitable for answering questions on a higher level, for example, gaining a general understanding of liver diagnostic strategies across the national level and identifying a few strategies from a pool of them. This can be helpful for future research studies to narrow it down to a specific population using a limited number of more cost-effective strategies. On the other hand, this data access challenge may offer research opportunities in developing predictive modeling tools that adapt information security and using fewer healthcare data to build more accurate predictive modeling tools.

Predictive modeling provides potential breakthroughs in healthcare and data analytics by addressing various needs across various healthcare domains. Robust and efficient predictive models can optimally leverage the digital health information generated every day or even every second to support different decision-making in a wide range of healthcare domains and to improve the patient outcomes and well-being of society.

APPENDIX A. SUPPLEMENTARY MATRIAL TO ESSAY 1

Introduction

Model structures and costs

A Bernoulli distribution was used to simulate the occurrence of EVs. A Bernoulli distribution is a probability distribution of a random variable taking on the value either 1 or 0, which corresponds to presence or absence, respectively, of some diagnosis. If a patient had EVs, then another Bernoulli distribution would be used to determine whether the EVs were small or large. If small varices were detected, the next EV screening would be two years later; whereas if large varices were found, patients would receive both beta-blockers and band ligation. The variceal bleeding in patients with varices and HCC development were determined every year with two other Bernoulli distributions.

Because the 5-year incidence of clinical outcomes (especially extrahepatic outcomes such cardiovascular, renal, or other organ cancers) and their mortality rates are unknown for this population, we did not include the costs of these events.

Fib-4 was calculated by summing costs of complete blood count and hepatic panel. Additionally, we included costs of liver biopsy-related complications. We did not consider costs associated with the treatment of HCC (except for liver transplantation or liver resection) or any liver-related decompensation.

Determination of the simulation replication (batch of cohort) number

As discussed in the Methods section, assessing outcomes of patients who receive a diagnosis of cirrhosis required consideration of uncertainties modeled with the Bernoulli distributions. The outcome results may be biased if only based on a small number of simulation replications (i.e., batches of the cohort). In other words, some individual strategy may perform better than others merely by chance. Hence, to obtain representative outcome results from the simulation, we determined the number of necessary simulation replications. The number of necessary replications

n was established when the absolute differences between the average per-patient total cost over the first *n* replications and the average per-patient total cost over the first *n* – 1 replications for all strategies tested were below our arbitrary threshold of $5 U.S. dollars. With the above criterion, we decided to run the simulation with 2 replications. In particular, after 2 replications for each strategy, we observed clear separation in the average total cost among the strategies, and the comparative results did not change as the replication number further increased (**Supplemental Figure A.1**).

Analysis

ICERs were calculated by dividing the incremental change in total cost by incremental change in diagnostic accuracy or percentage of correct diagnosis (cost strategy B – cost strategy A) / (accuracy strategy B – accuracy strategy A).

We conducted several replications of the microsimulation (i.e., joint samples from the Bernoulli distributions) to yield representative results for comparing the various diagnostic strategies, **see supplemental material (introduction) and supplemental Figure A.1**.

Accuracy calculation breakdown

We describe the calculation of the accuracy for single strategies and use MRE (**Supplemental Table A.6**) as an example. We had 40,000 failure cases after MRE. Then we did the confirmation test LB for the failure cases. The corresponding costs were added. Then accuracy was obtained by using 825,444 (TP+TN excluding confirmation results) divided by 1,000,000. The same logic applies to other single test strategies.

For computing accuracy of combo strategies, we use Fib-4 + VCTE (**Supplemental Table A.13**) as an example. After the first Fib-4 test, we had positive and indeterminate cases, and we continued the second VCTE test. After VCTE, putting both positive and indeterminate cases altogether, we had in total 30,473 failure cases. In the model, we further applied the confirmation LB test for these people. The corresponding costs were added. When calculating the accuracy, we used 892,615 (not including the confirmation results) divided by 1,000,000.

SUPPLEMENTAL TABLES

Table A.1. Prevalence of varices in NAFLD with advanced fibrosis.

Varices	Sample size	Diagnosis method	Variceal size	Compensated disease (CTP-A)	Prevalence in advanced fibrosis	Prevalence in cirrhosis
Nakamura S[1]	82	Liver biopsy	34.7% EV Small (48%) Large (52%)	72.2%	47.2% (GEV) 34.7% (EV)	N/A
Mendes F[2]	77 (subsample under GEV screening)	Liver biopsy	Small (59.6%) Large (40.4%)	100%	71.4% (GEV)	N/A

Abbreviations: **NAFLD**, nonalcoholic fatty liver disease; **GEV**, gastroesophageal varices; **EV**, esophageal varices. **CTP**, Child-Turcotte-Pugh score.

Table A.2. Studies reporting prevalence or annual incidence of HCC in NASH-related advanced fibrosis or cirrhosis.

Study	Sample Size	Mode of diagnosis	Prevalence	Annual Incidence of HCC	Prevalence of CTP B and C	Follow-up years
Sanyal[3]	149 NASH cirrhosis	Liver biopsy	6.7%	0.2%	78/152 (51%)	10
Bhala[4]	247 NASH-bridging fibrosis and cirrhosis	Liver biopsy	2.4%	0.3%	0%	7.1
Ascha[5]	195 NASH cirrhosis	Liver biopsy or cryptogenic fibrosis with metabolic syndrome	12.8%	2.6%	100% had a CTP ≥ 7 (B or C) or MELD ≥ 10	3.2
Berman[6]	156	CT	2.6%	6.1% at 5 years	N/A	4.7
Vilar-Gomez[7]	458 NASH-bridging fibrosis (149) and cirrhosis (299)	Liver biopsy	1.3% (bridging fibrosis) 13% (cirrhosis)	Bridging fibrosis- 0.2% Cirrhosis- 3.2%	0%	5.5

Abbreviations: HCC, hepatocellular carcinoma; **NASH**, nonalcoholic steatohepatitis; **CTP**, Child-Turcotte-Pugh score.

Table A.3. Bleeding- and HCC-related mortality.

Bleeding-related mortality - Overall population	
Pooled 6-monthly mortality[8]	16.3%
HCC-related mortality – Patients underwent HCC surveillance	
Pooled 6-monthly mortality[9, 10]	33.7%
HCC-related mortality – Overall population	
Pooled 5-year mortality[9]	72.1%

Abbreviations: HCC, hepatocellular carcinoma.

Table A.4. Cost of tests and treatments based on Medicare average.

Cost	Value ($)
Fib-4*	0
Liver biopsy (ultrasound guided)	1411
VCTE (Fibroscan)	150.34
MRI	544.18
US/AFP	159.61
EV (EGD facility)	773.19
EVL (EGD + banding)	1408.23
Beta-blockers	33.46 per month
Liver transplant	739100
Tumor/liver resection	21657.5

*Costs based on lab parameters (CBC + liver panel test).
Abbreviations: **VCTE**, vibration-controlled transient elastography; **MRI**, magnetic resonance imaging; **EV**, esophageal varices; **EGD**, esophagogastroduodenoscopy; **AFP**, alpha-fetoprotein; **US**, ultrasound.

Table A.5. Cost scenarios per procedure.

Procedure	Hospital based clinic (national average for facility fees)	Hospital based clinic (90^{th} percentile charges submitted)
Fib-4*	$16.76	$113
Liver biopsy (US guided)	$ 1,358.51	$ 5,657.20
MRI	$ 623.41	$ 4,804.19
CT	$ 558.21	$ 5,383.76
VCTE (Fibroscan)	$ 140.60	$ 589.50
US	$ 257.33	$ 1,337.03
EGD (diagnostic)	$ 820.59	$ 4,288.80
EGD with banding	$ 1,474.16	$ 5,544.33

*Costs based on lab parameters (CBC + liver panel test).

Abbreviations: **VCTE**, vibration-controlled transient elastography; **MRI**, magnetic resonance imaging; **EV**, esophageal varices; **EGD**, esophagogastroduodenoscopy; **AFP**, alpha-fetoprotein; **US**, ultrasound.

Table A.6. Accuracy and cost-effectiveness of different diagnostic strategies. A microsimulation analysis based on 100,000 NAFLD patients considering a cirrhosis prevalence of 4%.

Strategy	Number of correctly identified out of 4000 people with cirrhosis	Number of correctly excluded out of 96000 people without cirrhosis	Percentage of people correctly classified	Change in percentage of people correctly classified*	Cost per person ($)	Cost per correct diagnosis ($)	Mortality (bleeding)	Mortality (HCC)	Total mortality cases	Change in total mortality
No test	N/A	N/A	N/A	N/A	158	N/A	373	81	686	**Reference**
Fib-4+VCTE	1872	85659	87.5%	31.7%	1020	1170	316	60	613	-73
Fib-4+MRE	1863	88735	90.6%	34.7%	1108	1230	318	60	613	-73
Fib-4+LB	2152	93942	96.1%	40.2%	1433	1500	310	57	603	-83
VCTE+LB	3029	88073	91.1%	35.2%	1568	1733	283	47	568	-118
Fib-4	2224	53645	55.9%	**Reference**	1608	2892	306	56	599	-87
MRE+LB	3006	91366	94.4%	38.5%	1831	1950	284	47	569	-117
VCTE	3237	71975	75.2%	19.3%	1849	2474	276	44	559	-127
MRE	3221	79109	82.3%	26.5%	2066	2523	277	44	559	-127
LB	3720	91200	94.9%	39.1%	2765	2925	262	39	540	-146

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, Magnetic resonance elastography; **HCC**, hepatocellular carcinoma; **CC**, non-EV and non-HCC caused compensated cirrhosis
Strategies are listed in order of increasing costs.
* It represents difference between tests and the reference strategy (Fib-4).

Table A.7. Frontier table. One-way sensitivity analyses on the cost of each test.

Strategies	MRE			VCTE		LB		Fib4		Proportion of being on frontier
	Base	Mid	High	Low	High	Low	High	Mid	High	
Fib-4 + VCTE	**F***	**F**	**F**	**F**	-	**F**	**F**	**F**	**F**	89%
Fib-4 + MRE	**F**	**F**	-	**F**	**F**	**F**	**F**	**F**	**F**	89%
VCTE + LB	**-****	-	**F**	-	-	-	-	-	-	11%
Fib-4 + LB	**F**	**F**	-	**F**	**F**	**F**	**F**	**F**	**F**	89%
MRE + LB	-	-	-	-	-	-	**F**	-	-	11%

*"F" indicates that the corresponding strategy is on the frontier.
**"-" indicates that a strategy is dominated.
Note: All single test strategies are dominated, thus excluded from the above table.

Table A.8. Breakdown of test results in numbers for Fib-4 with cirrhosis prevalence 0.27%.

Test	Test result	Number of people	Number of people	True condition	Result type	Number of people
Fib-4	+	109,204	680,000	C	TP	698
				NC	FP	108,506
	-	570,796		C	FN	1,138
				NC	TN	569,658
	Indetermined (LB+)	16,760	320,000	C	TP	804
				NC	FP	15,957
	Indetermined (LB-)	303,240		C	FN	60
				NC	TN	303,179
	TP	FP	TN	FN	TP+TN	Accuracy
Including conformation with LB	1,501	124,463	872,837	1,199	874,338	87.4%
Excluding conformation with LB	698	108,506	569,658	1,138	570,355	83.9% (/680,000)*
						57% (/1,000,000)#

Abbreviations: VCTE, vibration controlled transient elastography; LB, liver biopsy; MRE, magnetic resonance elastography; C, cirrhosis; NC, no cirrhosis; TP, true positive, equivalent to CD, correctly diagnosed; FP, false positive, equivalent to MD, misdiagnosed; TN, true negative, equivalent to CR, correctly ruled out; FN, false negative, equivalent to UD, undiagnosed.

*Accuracy is calculated excluding confirmation LB test by TP+TN/number of people with test results from the test.

Accuracy is calculated excluding confirmation LB test by TP+TN/total number of people. This accuracy is used in our study.

Table A.9. Breakdown of test resutls in numbers for MRE with cirrhosis prevalence 0.27%.

Test	Test result	Number of people	Number of people	True condition	Result type	Number of people
MRE	+	136,111	960,000	C	TP	2074
				NC	FP	134037
	-	823,889		C	FN	518
				NC	TN	823371
	Failure (LB+)	2,095	40,000	C	TP	100
				NC	FP	1995
	Failure (LB-)	37,905		C	FN	8
				NC	TN	37897
	TP	FP	TN	FN	TP+TN	Accuracy
Including conformation with LB	2,174	136,032	861,268	526	863,442	86.3%
Excluding conformation with LB	2,074	134,037	823,371	518	825,444	86% (/960,000)*
						82.5% (/1,000,000)#

Abbreviations: VCTE, vibration controlled transient elastography; LB, liver biopsy; MRE, magnetic resonance elastography; C, cirrhosis; NC, no cirrhosis; TP, true positive, equivalent to CD, correctly diagnosed; FP, false positive, equivalent to MD, misdiagnosed; TN, true negative, equivalent to CR, correctly ruled out; FN, false negative, equivalent to UD, undiagnosed.

*Accuracy is calculated excluding confirmation LB test by TP+TN/number of people with test results from the test.

Accuracy is calculated excluding confirmation LB test by TP+TN/total number of people. This accuracy is used in our study.

Table A.10. Breakdown of test resutls in numbers for VCTE with cirrhosis prevalence 0.27%.

Test	Test result	Number of people	Number of people	True condition	Result type	Number of people
VCTE	+	178,040	929,000	C	TP	2,007
				NC	FP	176,033
	-	750,960		C	FN	502
				NC	TN	750,458
	Failure (LB+)	3,719	71,000	C	TP	178
				NC	FP	3,540
	Failure (LB-)	67,281		C	FN	13
				NC	TN	67,268
	TP	FP	TN	FN	TP+TN	Accuracy
Including conformation with LB	2,185	179,574	817,726	515	819,911	82%
Excluding conformation with LB	2,007	176,033	750,458	502	752,465	81% (/929,000)*
						75.2% (/1,000,000)#

Abbreviations: VCTE, vibration controlled transient elastography; LB, liver biopsy; MRE, magnetic resonance elastography; C, cirrhosis; NC, no cirrhosis; TP, true positive, equivalent to CD, correctly diagnosed; FP, false positive, equivalent to MD, misdiagnosed; TN, true negative, equivalent to CR, correctly ruled out; FN, false negative, equivalent to UD, undiagnosed.

*Accuracy is calculated excluding confirmation LB test by TP+TN/number of people with test results from the test.

Accuracy is calculated excluding confirmation LB test by TP+TN/total number of people. This accuracy is used in our study.

Table A.11. Breakdown of test resutls in numbers for LB with cirrhosis prevalence 0.27%.

Test	Test result	Number of people	Number of people	True condition	Result type	Number of people
LB	+	52,376	1,000,000	C	TP	2,511
				NC	FP	49,865
	-	947,624		C	FN	189
				NC	TN	947,435
	TP	FP	TN	FN	TP+TN	Accuracy
	2,511	49,865	947,435	189	949,946	95%

Abbreviations: VCTE, vibration controlled transient elastography; LB, liver biopsy; MRE, magnetic resonance elastography; C, cirrhosis; NC, no cirrhosis; TP, true positive, equivalent to CD, correctly diagnosed; FP, false positive, equivalent to MD, misdiagnosed; TN, true negative, equivalent to CR, correctly ruled out; FN, false negative, equivalent to UD, undiagnosed.

Table A.12. Breakdown of test resutls in numbers for Fib-4+MRE with cirrhosis prevalence 0.27%.

1st test	1st test result	Number of people	2nd test	2nd test result	Number of people	True condition	Result type	Number of people
Fib4	+	109,204	MRE	+	15,119	C	TP	536
						NC	FP	14,583
				-	89,717	C	FN	134
						NC	TN	89,583
				Failure (LB+)	243	C	TP	26
						NC	FP	217
				Failure (LB-)	4,125	C	FN	2
						NC	TN	4,123
	Indetermined	320,000		+	43,555	C	TP	664
						NC	FP	42,892
				-	263,645	C	FN	166
						NC	TN	263,479
				Failure (LB+)	670	C	TP	32
						NC	FP	638
				Failure (LB-)	12,130	C	FN	2
						NC	TN	12,127
	-	570,796	/	/	/	C	FN	1,138
						NC	TN	569,658

Summary	Total failure	Total non-failure
Number of people	17,168	982,832

	TP	FP	TN	FN	TP+TN	Accuracy
Including conformation with LB	1,257	58,330	938,970	940,227		94.0% 94% (/982,832)*
Excluding conformation with LB	1,199	57,475	922,719	923,919		92.4% (/1,000,000)#

Abbreviations: VCTE, vibration controlled transient elastography; LB, liver biopsy; MRE, magnetic resonance elastography; C, cirrhosis; NC, no cirrhosis; TP, true positive, equivalent to CD, correctly diagnosed; FP, false positive, equivalent to MD, misdiagnosed; TN, true negative, equivalent to CR, correctly ruled out; FN, false negative, equivalent to UD, undiagnosed.
*Accuracy is calculated excluding confirmation LB test by TP+TN/number of people with test results from the test.
Accuracy is calculated excluding confirmation LB test by TP+TN/total number of people. This accuracy is used in our study.

Table A.13. Breakdown of test resutls in numbers for Fib-4+VCTE with cirrhosis prevalence 0.27%.

1st test	1st test result	Number of people	2nd test	2nd test result	Number of people	True condition	Result type	Number of people
Fib4	+	109,204	VCTE	+	19671	C	TP	519
						NC	FP	19,152
				-	81779	C	FN	130
						NC	TN	81,650
				Failure (LB+)	431	C	TP	46
						NC	FP	385
				Failure (LB-)	7322	C	FN	3
						NC	TN	7,319
	Indetermined	320,000		+	56973	C	TP	642
						NC	FP	56,331
				-	24030	C	FN	161
						NC	TN	240,147
				Failure (LB+)	1190	C	TP	57
						NC	FP	1,133
				Failure (LB-)	21530	C	FN	4
						NC	TN	21,526
	-	570,796	/	/	/	C	FN	1,138
						NC	TN	569,658

Summary	Total failure	Total non-failure
Number of people	30,473	969,527

	TP	FP	TN	FN	TP+TN	Accuracy
Including conformation with LB	1,264	77,001	920,299	1,436	921,562	92.2% 92.1% (/969,527)*
Excluding conformation with LB	1,161	75,483	891,454	1,428	892,615	89.3% (/1,000,000)#

1st test	1st test result	Number of people	2nd test	2nd test result	Number of people	True condition	Result type	Number of people
Fib4	+	109,204	LB	+	6,074	C	TP	649
						NC	FP	5,425
				-	103,130	C	FN	49
						NC	TN	103,081
	Indetermined	320,000		+	16760	C	TP	804
						NC	FP	15,957
				-	303240	C	FN	60
						NC	TN	303,179
	-	570,796	/	/	/	C	FN	1,138
						NC	TN	569,658

	TP	FP	TN	FN	TP+TN	Accuracy
Including conformation with LB	1,452	21,382	975,918	1,248	977,370	97.7%

Abbreviations: VCTE, vibration controlled transient elastography; LB, liver biopsy; MRE, magnetic resonance elastography; C, cirrhosis; NC, no cirrhosis; TP, true positive, equivalent to CD, correctly diagnosed; FP, false positive, equivalent to MD, misdiagnosed; TN, true negative, equivalent to CR, correctly ruled out; FN, false negative, equivalent to UD, undiagnosed.
*Accuracy is calculated excluding confirmation LB test by TP+TN/number of people with test results from the test.
Accuracy is calculated excluding confirmation LB test by TP+TN/total number of people. This accuracy is used in our study.

Table A.14. Breakdown of test resutls in numbers for Fib-4+LB with cirrhosis prevalence 0.27%.

1st test	1st test result	Number of people	2nd test	2nd test result	Number of people	True condition	Result type	Number of people
MRE	+	136,111	LB	+	8,630	C	TP	1,928
						NC	FP	6,702
				-	127,480	C	FN	145
						NC	TN	127,335
	-	823,889	/	/	/	C	FN	518
						NC	TN	823,371
	Failure (LB+)	2,095	/	/	/	C	TP	100
						NC	FP	1,995
	Failure (LB-)	37,905	/	/	/	C	FN	8
						NC	TN	37,897

Summary	Total failure	Total non-failure
Number of people	40,000	960,000

	TP	FP	TN	FN	TP+TN	Accuracy
Including conformation with LB	2,029	8,696	988,604	671	990,632	99.1%
Excluding conformation with LB	1,928	6,702	950,706	664	952,635	99.2% (/960,000)* 95.3% (/1,000,000)#

Abbreviations: VCTE, vibration controlled transient elastography; LB, liver biopsy; MRE, magnetic resonance elastography; C, cirrhosis; NC, no cirrhosis; TP, true positive, equivalent to CD, correctly diagnosed; FP, false positive, equivalent to MD, misdiagnosed; TN, true negative, equivalent to CR, correctly ruled out; FN, false negative, equivalent to UD, undiagnosed.

Table A.15. Breakdown of test resutls in numbers for MRE+LB with cirrhosis prevalence 0.27%.

1st test	1st test result	Number of people	2nd test	2nd test result	Number of people	True condition	Result type	Number of people
VCTE	+	178,040	LB	+	8,630	C	TP	1,866
						NC	FP	8,802
				-	127,480	C	FN	140
						NC	TN	167,232
	-	750,960	/	/	/	C	FN	502
						NC	TN	750,458
	Failure (LB+)	3,719	/	/	/	C	TP	178
						NC	FP	3,540
	Failure (LB-)	67,281	/	/	/	C	FN	13
						NC	TN	67,268

Summary	Total failure	Total non-failure
Number of people	71,000	929,000

	TP	FP	TN	FN	TP+TN	Accuracy
Including conformation with LB	2,044	12,342	984,958	656	987,002	98.7%
Excluding conformation with LB	1,866	8,802	917,690	642	919,556	99% (/929,000)* 92% (/1,000,000)#

Abbreviations: VCTE, vibration controlled transient elastography; LB, liver biopsy; MRE, magnetic resonance elastography; C, cirrhosis; NC, no cirrhosis; TP, true positive, equivalent to CD, correctly diagnosed; FP, false positive, equivalent to MD, misdiagnosed; TN, true negative, equivalent to CR, correctly ruled out; FN, false negative, equivalent to UD, undiagnosed.
*Accuracy is calculated excluding confirmation LB test by TP+TN/number of people with test results from the test.
Accuracy is calculated excluding confirmation LB test by TP+TN/total number of people. This accuracy is used in our study.

Table A.16. Breakdown of test resutls in numbers for VCTE+LB with cirrhosis prevalence 0.27%.

1st test	1st test result	Number of people	2nd test	2nd test result	Number of people	True condition	Result type	Number of people
VCTE	+	178,040	LB	+	8,630	C	TP	1,866
						NC	FP	8,802
				-	127,480	C	FN	140
						NC	TN	167,232
	-	750,960	/	/	/	C	FN	502
						NC	TN	750,458
	Failure (LB+)	3,719	/	/	/	C	TP	178
						NC	FP	3,540
	Failure (LB-)	67,281	/	/	/	C	FN	13
						NC	TN	67,268

Summary	Total failure	Total non-failure
Number of people	71,000	929,000

	TP	FP	TN	FN	TP+TN	Accuracy
Including conformation with LB	2,044	12,342	984,958	656	987,002	98.7%
Excluding conformation with LB	1,866	8,802	917,690	642	919,556	99% (/929,000)* 92% (/1,000,000)#

Abbreviations: VCTE, vibration controlled transient elastography; LB, liver biopsy; MRE, magnetic resonance elastography; C, cirrhosis; NC, no cirrhosis; TP, true positive, equivalent to CD, correctly diagnosed; FP, false positive, equivalent to MD, misdiagnosed; TN, true negative, equivalent to CR, correctly ruled out; FN, false negative, equivalent to UD, undiagnosed.
*Accuracy is calculated excluding confirmation LB test by TP+TN/number of people with test results from the test.
Accuracy is calculated excluding confirmation LB test by TP+TN/total number of people. This accuracy is used in our study.

Table A.17. The incremental cost-effectiveness ratios for each combination strategy based on different cirrhosis prevalence settings.*

Diagnostic strategy	Cirrhosis prevalence								
	----------0.27%----------			**----------2%----------**			**----------4%----------**		
	Diagnostic accuracy	Cost per person ($)	ICERs	Diagnostic accuracy	Cost per person ($)	ICERs	Diagnostic accuracy	Cost per person ($)	ICERs
Fib-4 + VCTE	89.3%	401	**Least costly**	88.5%	687	**Least costly**	87.5%	1020	**Least costly**
Fib-4 + MRE	92.4%	491	2864	91.6%	778	2911	90.6%	1108	2879
Fib-4 + LB	97.7%	729	4454	97.0%	1055	5114	96.1%	1433	5903

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography; **ICER**, the incremental cost-effectiveness ratio.
*Strategies not listed are dominated and, thus, excluded.
†Fib-4+LB was dominated by the next strategy (VCTE+LB).

Table A.18. Cost per death prevented (CPDP) with cirrhosis prevalence 0.27%.

Diagnostic strategy	Cirrhosis prevalence ----------**0.27%**----------		
	Mortality	Cost per person ($)	CPDP
No test	46	10	**Least costly**
Fib-4+VCTE	42	401	93
Fib-4+MRE	42	491	Dominated
VCTE+LB	40	612	92
Fib-4+LB	42	729	Dominated
MRE+LB	40	888	Dominated
VCTE	39	900	460
Fib-4	41	908	Dominated
MRE	39	1109	Dominated
LB	37	1663	372

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography; **CPDP**, the cost per death prevented.
CPCD is obtained by (cost of A – cost of B)/(Mortality of B – Mortality of A), where A is the current least costly strategy, and B is the next least costly strategy. The test strategies are arranged in cost ascending order, and a test strategy is dominated if CPDP is negative.

Table A.19. Cost per death prevented (CPDP)

Diagnostic strategy	Cirrhosis prevalence ----------2%----------		
	Mortality	Cost per person ($)	CPDP
No test	344	77	**Least costly**
Fib-4+VCTE	307	687	16
Fib-4+MRE	307	778	Dominated
VCTE+LB	285	1055	13
Fib-4+LB	302	1055	Dominated
Fib-4	300	1231	Dominated
MRE+LB	285	1325	Dominated
VCTE	279	1341	48
MRE	280	1553	Dominated
LB	270	2177	93

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography; **CPDP**, the cost per death prevented.

CPCD is obtained by (cost of A – cost of B)/(Mortality of B – Mortality of A), where A is the current least costly strategy, and B is the next least costly strategy. The test strategies are arranged in cost ascending order, and a test strategy is dominated if CPDP is negative.

Table A.20. Cost per death prevented (CPDP)

Diagnostic strategy	Cirrhosis prevalence ----------4%----------		
	Mortality	Cost per person ($)	CPDP
No test	686	158	**Least costly**
Fib-4+VCTE	613	1020	12
Fib-4+MRE	613	1108	Dominated
Fib-4+LB	603	1433	41
VCTE+LB	568	1568	4
Fib-4	599	1608	Dominated
MRE+LB	569	1831	Dominated
VCTE	559	1849	31
MRE	559	2066	Dominated
LB	540	2765	48

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography; **CPDP**, the cost per death prevented.

CPCD is obtained by (cost of A – cost of B)/(Mortality of B – Mortality of A), where A is the current least costly strategy, and B is the next least costly strategy. The test strategies are arranged in cost ascending order, and a test strategy is dominated if CPDP is negative.

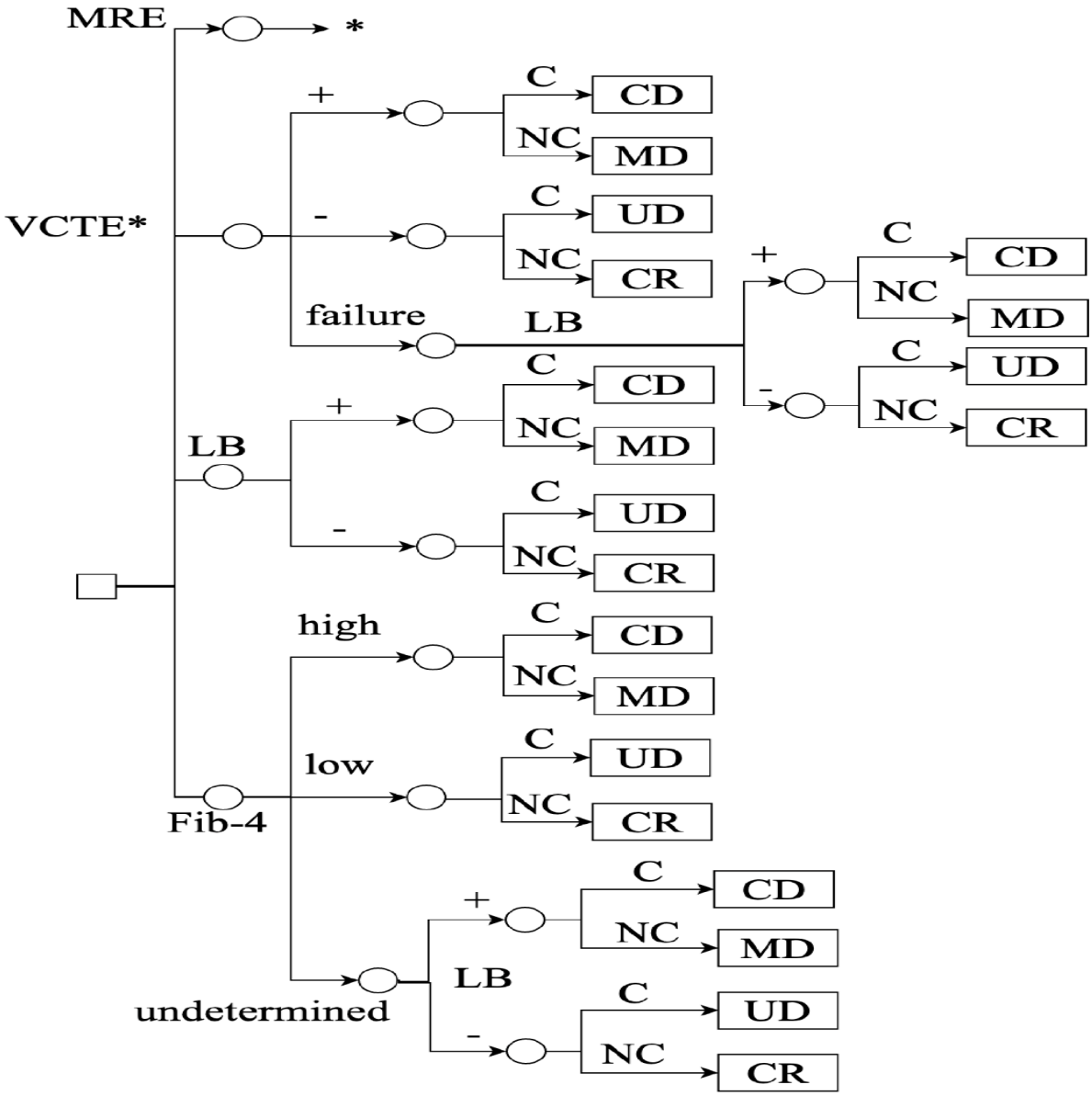


Figure A.1. Average total cost associated with increasing number of simulation replications .

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography.
Lines represent mean total cost for each strategy among all replications.

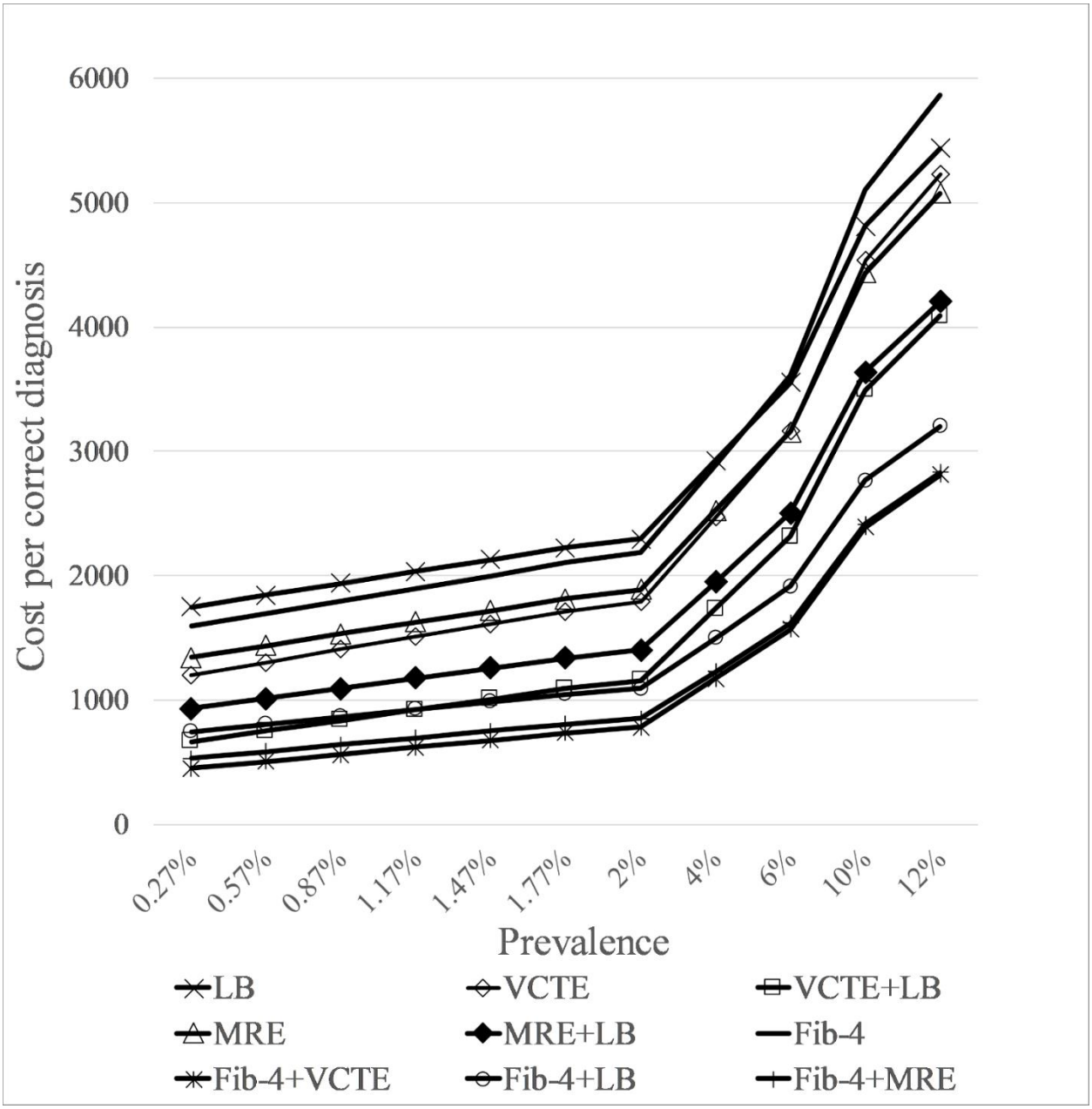
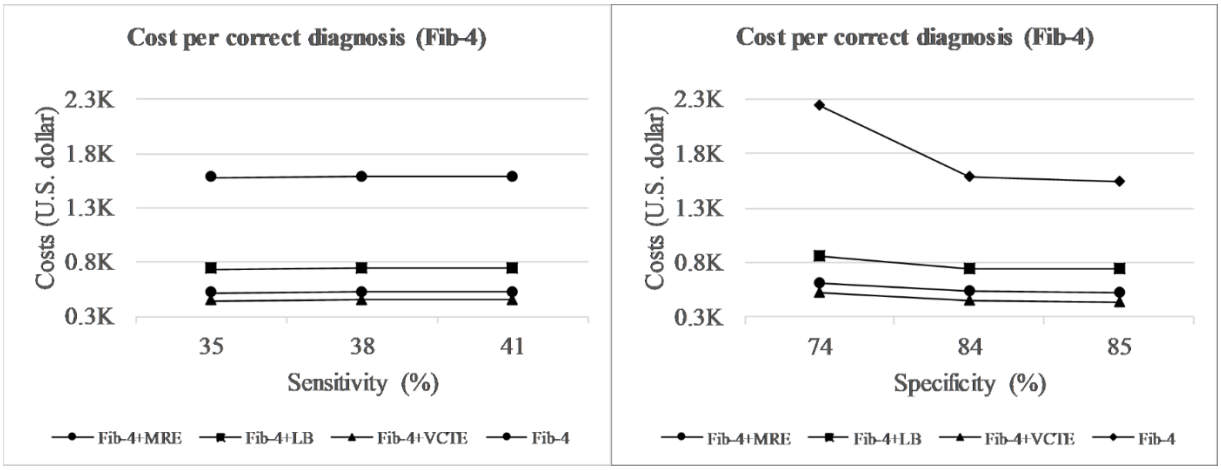
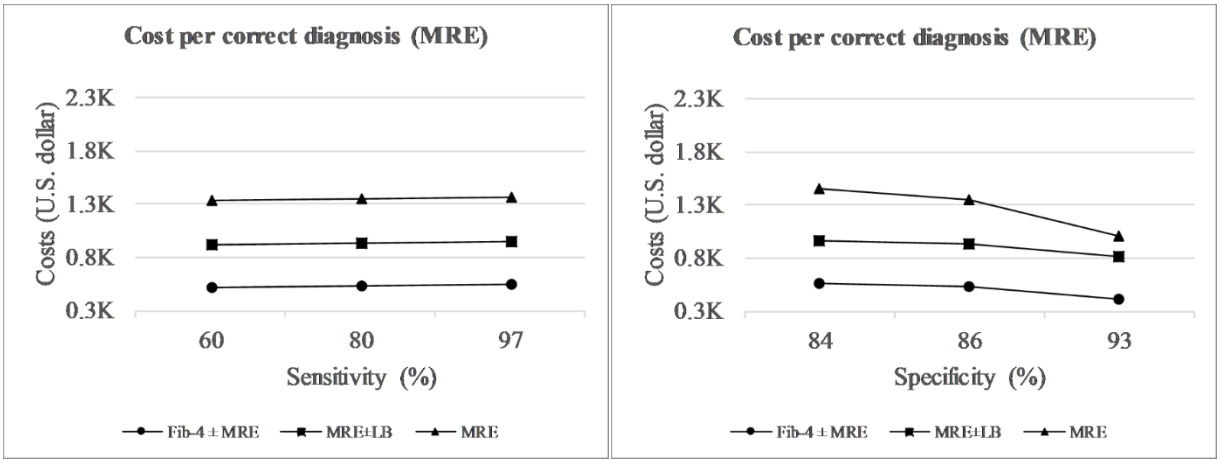


Figure A.2. Cost per correct diagnosis of cirrhosis vs. diagnostic accuracy for each diagnostic strategy. Extended analysis considering cirrhosis prevalence between 0.27 and 12%.

(A) Fib-4 index sensitivity and specificity



(B) MRE sensitivity and specificity



(C) VCTE sensitivity and specificity

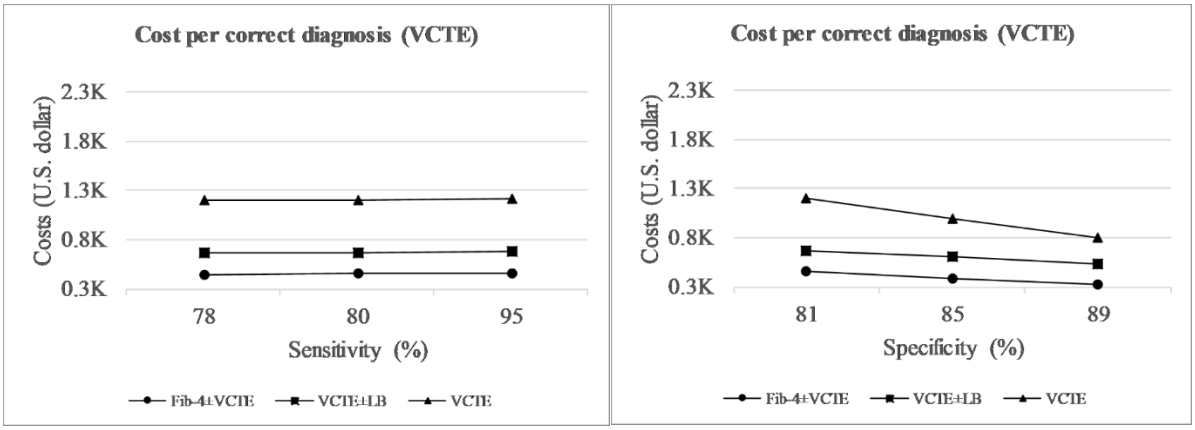


Figure A.3. Sensitivity analysis on baseline test characteristics. Analysis based on Medicare average.

(D) Liver biopsy sensitivity and specificity

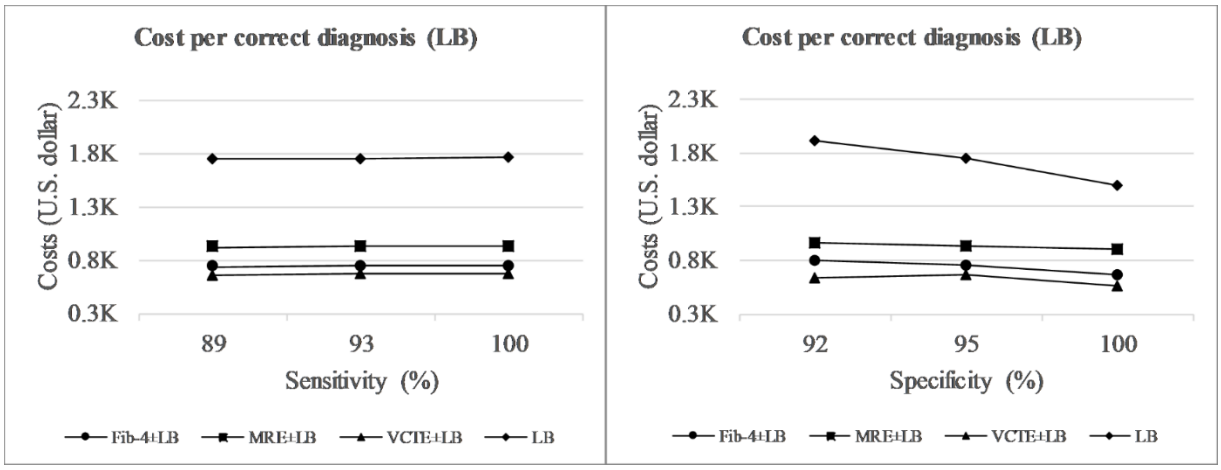


Figure A.3. continued

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography.

Note: The increase in sensitivity would leads to higher cost per diagnoses (i.e., when more people are correctly diagnosed with cirrhosis, more people will have to be involved in the downstream activity and as a result higher cost. On the contrary, the increase in the specificity would leads to increase in the number of people who will not need to participate in the downstream events with lower cost.

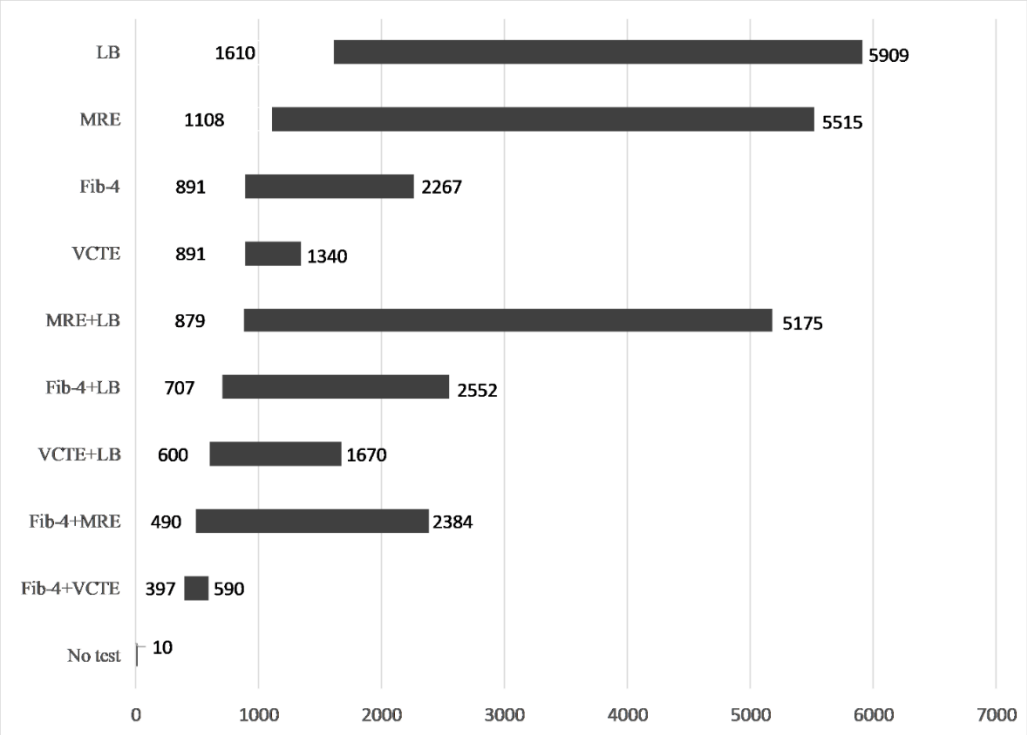


Figure A.4. Tornado plot for one-way sensitivty analysis on costs of Fib-4, MRE, VCTE and LB.

Abbreviations: VCTE, vibration controlled transient elastography; **LB**, liver biopsy; **MRE**, magnetic resonance elastography.

REFERENCES to Appendix A

1. Nakamura S, Konishi H, Kishino M, et al. Prevalence of esophagogastric varices in patients with non-alcoholic steatohepatitis. Hepatol Res 2008;38:572-9.
2. Mendes FD, Suzuki A, Sanderson SO, et al. Prevalence and indicators of portal hypertension in patients with nonalcoholic fatty liver disease. Clin Gastroenterol Hepatol 2012;10:1028-33 e2.
3. Sanyal AJ, Banas C, Sargeant C, et al. Similarities and differences in outcomes of cirrhosis due to nonalcoholic steatohepatitis and hepatitis C. Hepatology 2006;43:682-689.
4. Bhala N, Angulo P, van der Poorten D, et al. The Natural History of Nonalcoholic Fatty Liver Disease With Advanced Fibrosis or Cirrhosis: An International Collaborative Study. Hepatology 2011;54:1208-1216.
5. Ascha MS, Hanouneh IA, Lopez R, et al. The Incidence and Risk Factors of Hepatocellular Carcinoma in Patients with Nonalcoholic Steatohepatitis. Hepatology 2010;51:1972-1978.
6. Berman K, Tandra S, Vuppalanchi R, et al. Hepatic and Extrahepatic Cancer in Cirrhosis: A Longitudinal Cohort Study (vol 106, pg 899, 2011). American Journal of Gastroenterology 2011;106:1405-1405.
7. Vilar-Gomez E, Calzadilla-Bertot L, Wong VWS, et al. Fibrosis Severity as a Determinant of Cause-Specific Mortality in Patients With Advanced Nonalcoholic Fatty Liver Disease: A Multi-National Cohort Study. Gastroenterology 2018;155:443-57 e17.
8. Stevenson M, Lloyd-Jones M, Morgan MY, et al. Non-invasive diagnostic assessment tools for the detection of liver fibrosis in patients with suspected alcohol-related liver disease: a systematic review and economic evaluation. Health Technol Assess 2012;16:1-174.
9. Heimbach JK, Kulik LM, Finn RS, et al. AASLD guidelines for the treatment of hepatocellular carcinoma. Hepatology 2018;67:358-380.
10. Dakin H, Bentley A, Dusheiko G. Cost-utility analysis of tenofovir disoproxil fumarate in the treatment of chronic hepatitis B. Value Health 2010;13:922-33.

APPENDIX B. SUPPLEMENTARY MATERIAL TO ESSAY 2

Model Overview

We developed a microsimulation model of CVD for NYC adults based on the well-established CVD Policy Model. As a microsimulation model, our model can calculate CVD risks at the individual level based on time-varying risk factors. Baseline individual risk factors were estimated based on the NYC Health and Nutrition Examination Surveys (HANES) database of adults age 18+ years. The NYC HANES population was then matched to the Columbia University-National Heart Lung and Blood Institute Pooled Cohorts Dataset[89] to determine the trajectories of individual risk factors.

We used logistic functions to calculate the probabilities of developing CHD or stroke event in a certain year. The co-variates used to calculate CHD and stroke risks include age, race, body mass index (BMI), former smoker status, current smoker status, cigarettes per day, systolic blood pressure, diabetes status (fasting glucose ≥126 mg/dl [7.0 mmol/L] or taking anti-diabetes medications), high-density lipoprotein cholesterol, low-density lipoprotein cholesterol, and estimated glomerular filtration rate. Detailed model equations and associated parameter values are presented in Table B.1. To achieve representativeness of the NYC population in our modeling study, 10,000 individuals were randomly sampled with replacement from the NYC HANES population using sampling weights. Descriptive statistics for the sample and simulated populations are presented in Table B.3.

Our model includes the following health states: no CVD events (healthy), one or more CHD events (CHD), one or more stroke events, both CHD and stroke events, non-CVD related death, and CVD related death. The CHD heath state consists of the first CHD event of a year, and a recurrent CHD event of the same year. The stroke heath state consists of the first stroke event of a year, and a recurrent stroke event of the same year. When the simulation runs, all individuals start with the healthy state, and each individual experiences at most two CVD related events per year. At any given time, an individual can transition from the healthy state to a CVD state. An individual who has experienced any CVD event cannot return to the healthy state but has a risk of CVD-related death or a recurrent CHD or stroke event in the same cycle/year. Figure B.1 provides a schematic of the model structure. The simulation can be configured for any duration of time period of interest (e.g., 10 years, lifetime) with a closed population design.

For each policy scenario as well as status quo, the model generates predicted changes in the probability of each health outcome at the individual level. The model also tracks the health and economic impacts of each policy (e.g., SSB taxes, FV subsidies, and financing FV subsidies with an SSB tax. More specifically, each health state has an attributed quality of life utility (i.e., an overall assessment of well-being on a scale from 0 [death] to 1 [perfect health]) and cost. The model then estimates costs and QALYs based on each individual's health history including CHD, stroke, and death events.

Model Inputs

Probability of first-ever incident CVD and non-CVD related death

The annual first-ever incident probability is the one-year probability of healthy individuals dying or experiencing a CHD or stroke event. We used logistic function to model the probabilities of first-ever incident CVD and non-CVD related death (equation 1) as:

$$P_{k,i} = \frac{\exp(\alpha + \boldsymbol{\beta X})}{1 + \exp(\alpha + \boldsymbol{\beta X})}. \quad (1)$$

In this equation, $P_{k,i}$ denotes the annual probability of healthy individual i experiencing first CVD event k. The parameter α represents the underlying rate for event *k* in the pooled cohort population. The term $\boldsymbol{X}$ is a vector of CVD risk factors including age, sex, race, BMI, smoking status (former, current, cigarettes per day), systolic blood pressure, low-density lipoprotein cholesterol (LDL-c), high-density lipoprotein cholesterol (HDL-c), diabetes, and estimated glomerular filtration rate. We also included interaction terms between age and each of the 5 factors, including current smoker, systolic blood pressure, diabetes, low-density lipoprotein cholesterol (LDL-c), and high-density lipoprotein cholesterol (HDL-c). Similarly, the first-ever incidence of stroke varies with all risk factors exposures above except BMI, former smoking status, and cigarettes per day. The associated interaction terms are between age and each of the four factors including race, current smoker, systolic blood pressure, and diabetes. Non-CVD related mortality rate varies with all risk factors exposures above from CHD incidence except LDL-c and HDL-c. We also incorporated BMI^2, interaction terms that are between age and each of the three factors including race, BMI^2 and diabetes.

Probability of recurrent CVD events

The transitions among the population living with CVD states include: recurrent CHD event within a year of a prior occurrence, recurrent CHD event after a year of a prior occurrence, first/second stroke event per year after stroke within 1 year, stroke after CHD, first/second CHD after stroke within 10 years, and CHD proceeding stroke after 10 years. We derived the estimates of the above transition probabilities from previous empirical studies based on community-dwelling patients living with chronic CVD or on hospital-based CVD case registries, stratified by age and gender (see Table B.8).

Probability of survival to 30 days after an acute CVD event

The model incorporates the 30-day case fatality rates in individuals experiencing CHD and stroke events, stratified by age and sex (Table B.2). The 30-day case fatality rate for CHD events differs between first-ever incident and recurrent CHD events. For stroke, we assumed the same 30-day case fatality rates for first-ever incident and recurrent events (see Table B.2).

Probability of at home fatality after at least one CHD event

We incorporated calibrated estimates stratified by age and gender (See Table B.2).

CVD background direct medical costs and utilities

Health-related quality of life, CVD and background direct medical costs specifications can be found in a previous study.[89] We derived our estimates from multiple data sources (see Table B.7 for costs and Table B.8 for utilities).

Policy costs for SSB taxes and F&V subsidy

Multiple estimates for the cost of implementing an SSB tax exist, with most representing the cost as a function of tax revenue. For example, the Brookings Institution has gauged that the average sales tax has a government collection cost, and taxpayer compliance cost of 2% – 5%, while value-added taxes have similar cost of 3- 5%.[146] The Organization for Economic Cooperation and Development (OECD) evaluated the cost collection ratio for administrative costs (in the US) to be 0.47% – 0.66%.[147] In the specific case of the SSB tax implementation in the city of Berkeley, the firm contracted to administer the SSB tax charged 2% of tax revenues. Thus, we included policy implementation costs in the model as 2% of the SSB tax revenue collected, in line with our literature review and with current empirical estimates for cost of administration for such a tax.

The FV subsidy costs included the administrative costs, and the subsidy costs. To estimate the administrative costs, we considered sources from SNAP and Medicaid given similarities to the proposed intervention design and population of interest.[89] We assumed 20% of the total subsidy costs to be the administrative costs in the first year, conservatively based on the administrative costs (i.e., the percentage of total benefits) for the SNAP program in the first year when the EBT system was introduced.[148] This amount would include many other existing administrative costs of SNAP beyond the set-up of the EBT system. No data are available on the incremental administrative costs of the EBT system in SNAP after the first year.[89] The Healthy Incentives Program (HIP) trial within SNAP found that most of the implementation costs were one-time costs.[149] The HIP report estimated that the administrative implementation costs in the first year would be 6.2% of total subsidy costs in the first full year of implementation, and administrative implementation costs would likely decrease after the first year. Thus, we assumed 5% for administrative costs after the first year.

The formula to calculate policy costs are as follows:
SSB tax policy cost = - SSB intake (per serving) after intervention × Per serving SSB price changes after intervention.
FV subsidy cost = FV intake (per serving) after intervention × Per serving FV price changes after intervention.

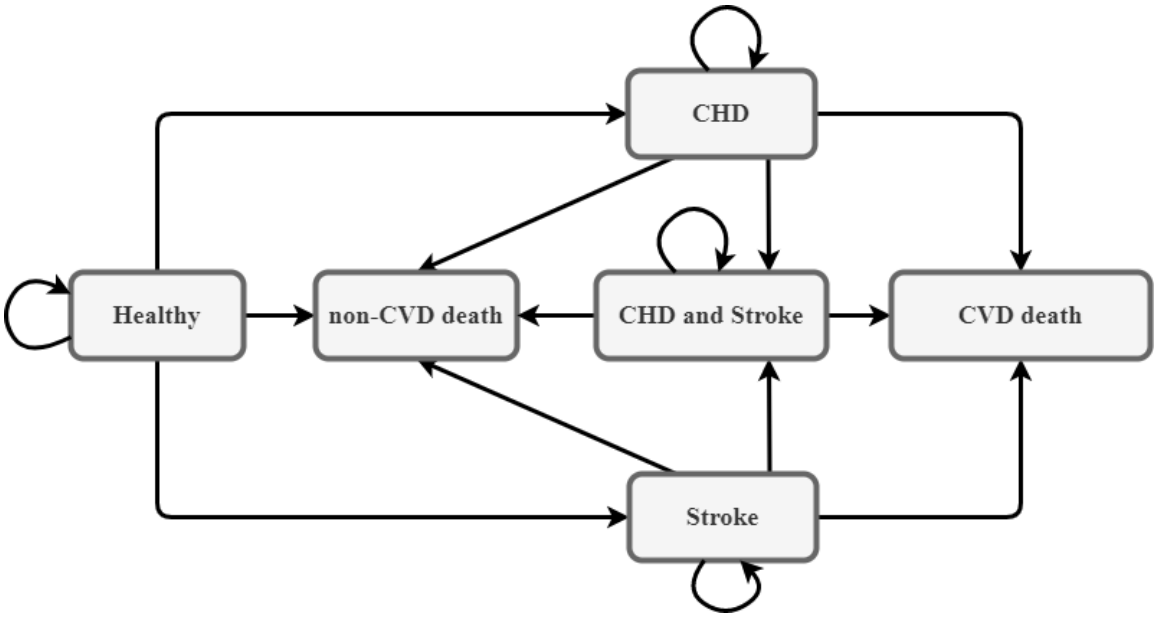


Figure B.1. Model Schematic

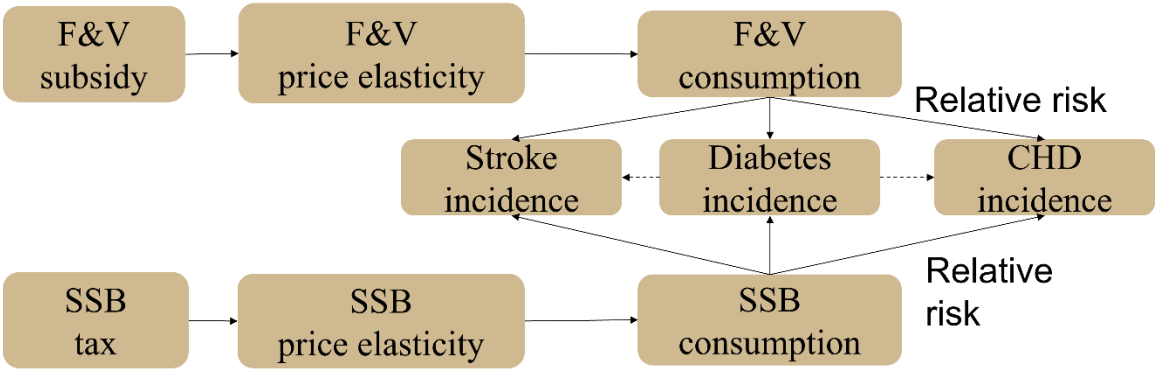
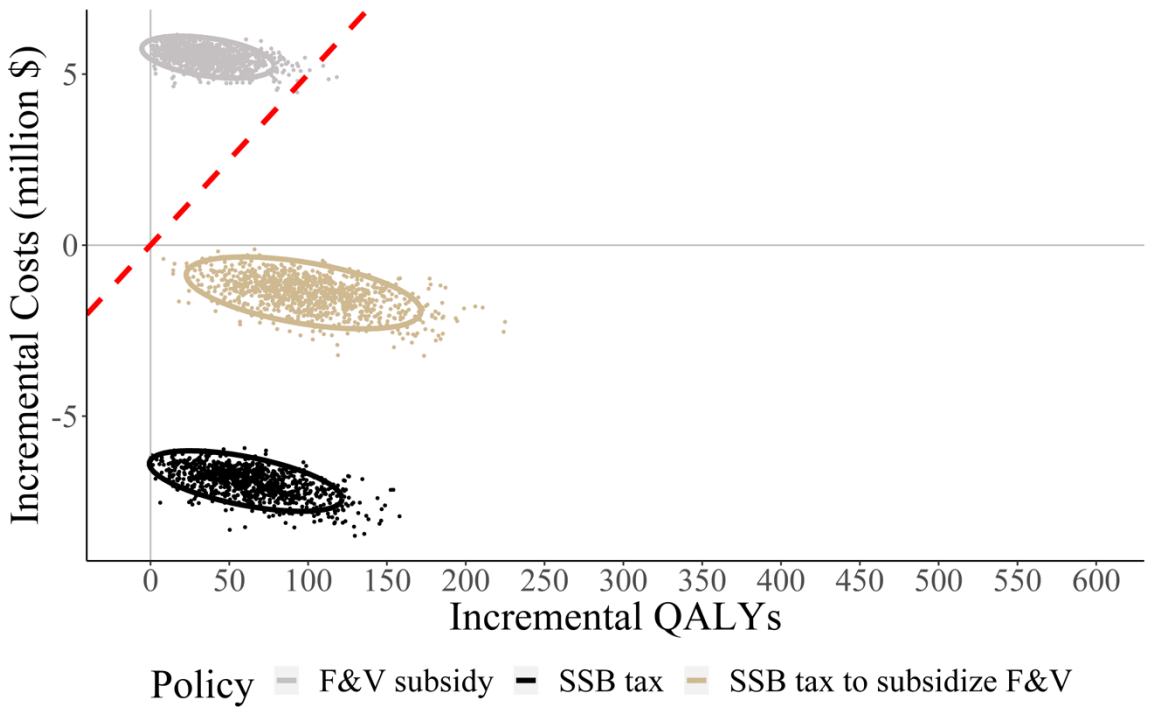
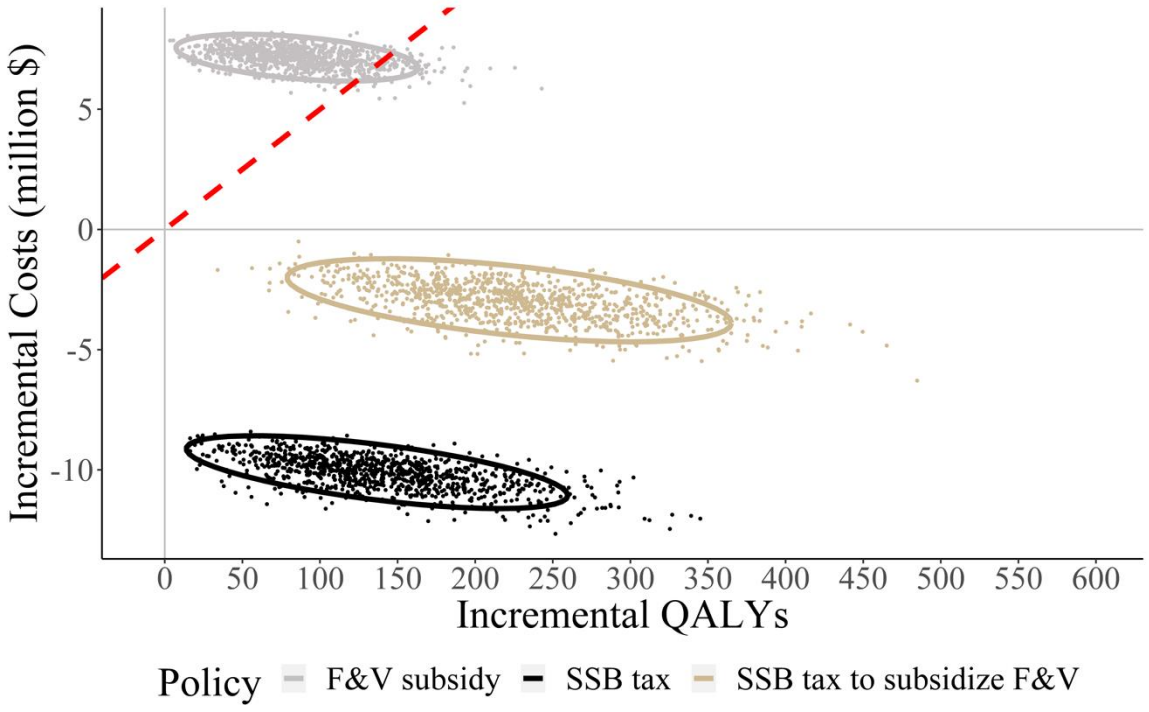


Figure B.2. Logic Pathway of Linking SSB Taxes and F&V Subsidy to Changes in CHD and Stroke Incidence

The logic pathway was described in the main text method section, Effects of Fruit, Vegetable and SSB Intake Changes on Cardiometabolic Risk. The dotted lines represent that diabetes was treated as a CVD risk factor to predict first-ever CVD cases.

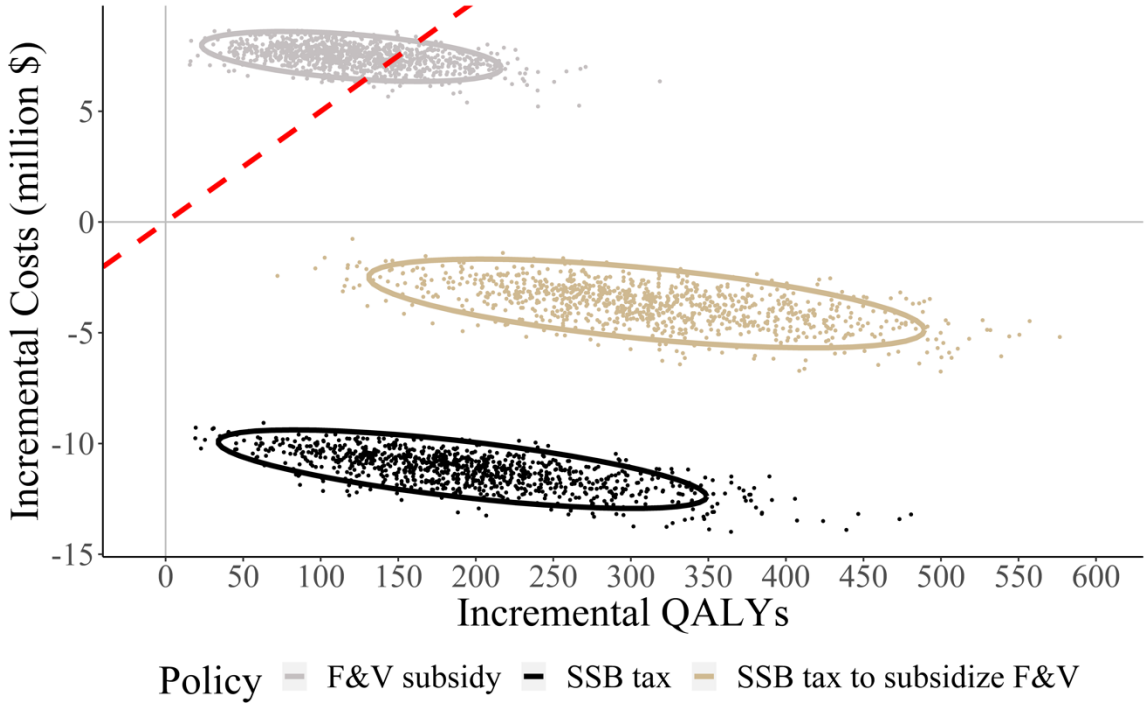


20 years



40 years

Figure B.3a. ICER Plot Over 20 Years VS. 40 Years vs. Lifetime from Societal Perspective



Lifetime

Figure B.3a. continued

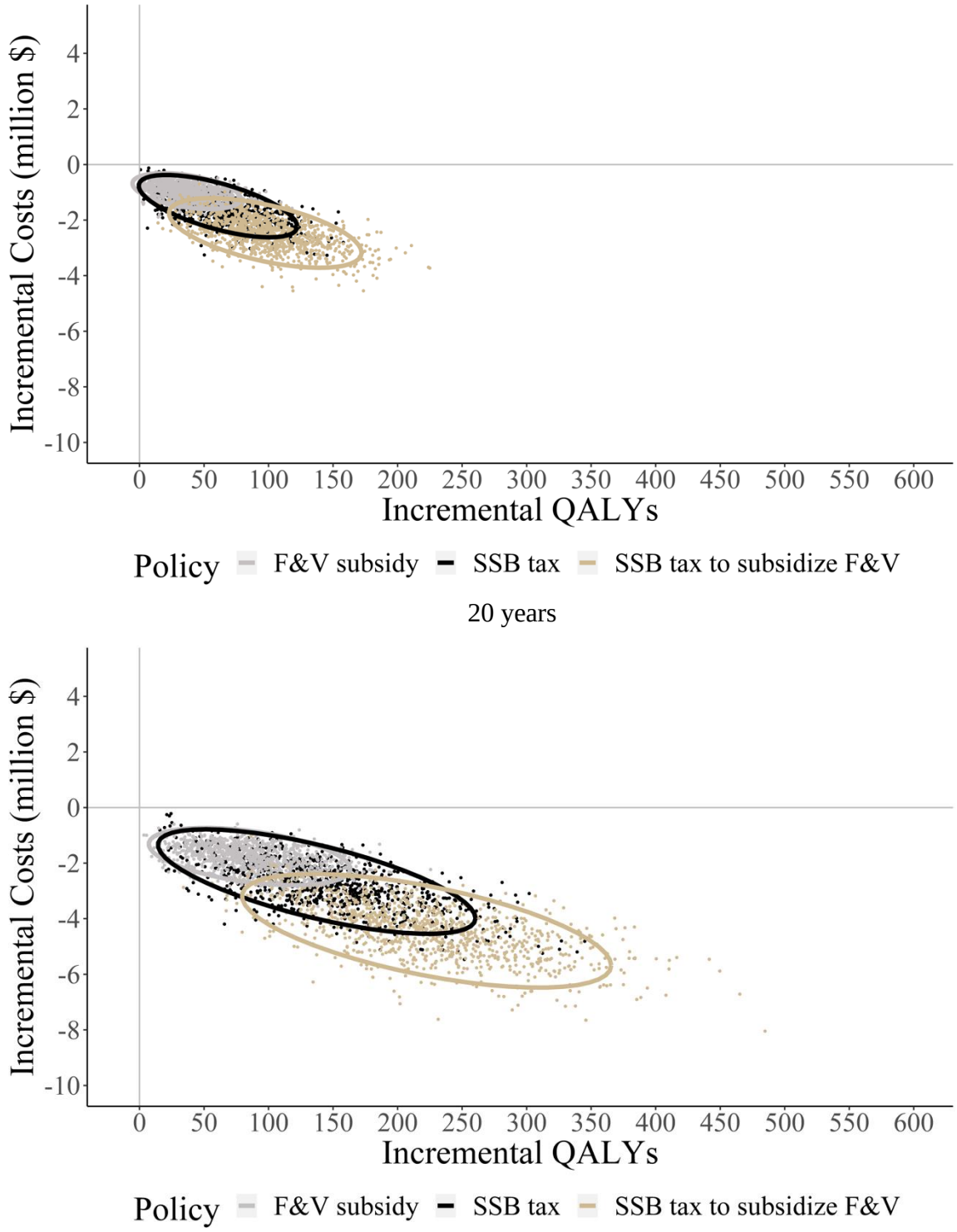
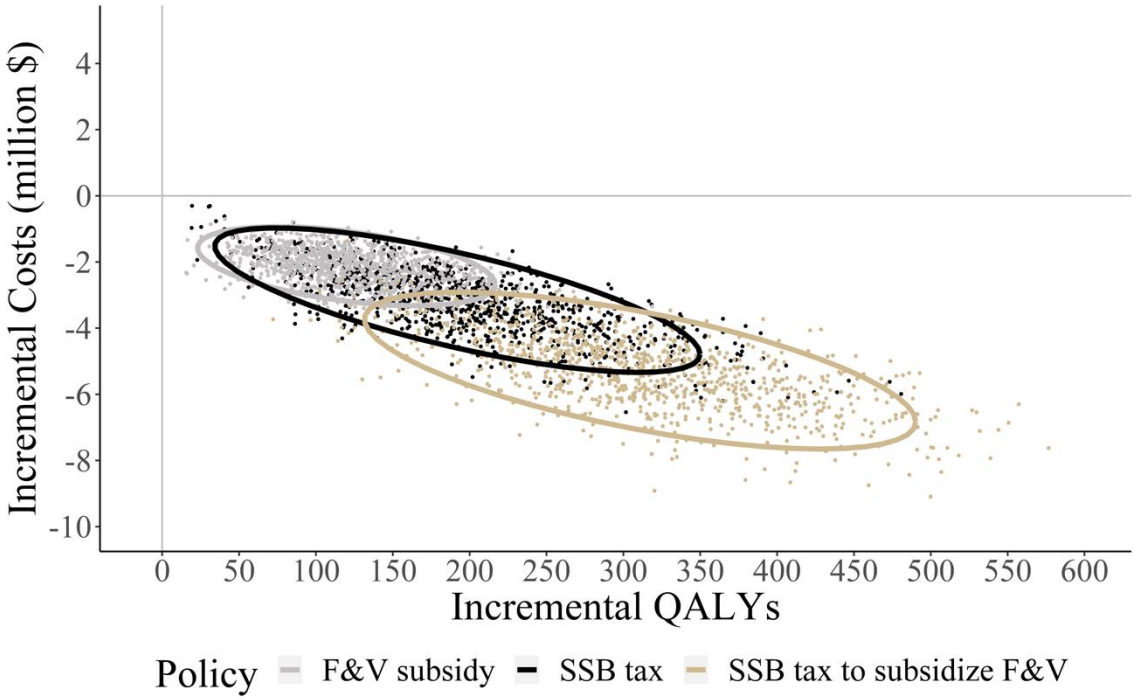


Figure B.3b. ICER Plot Over 20 Years VS. 40 Years VS. Lifetime from Healthcare Perspective



Lifetime

Figure B.3b. continued

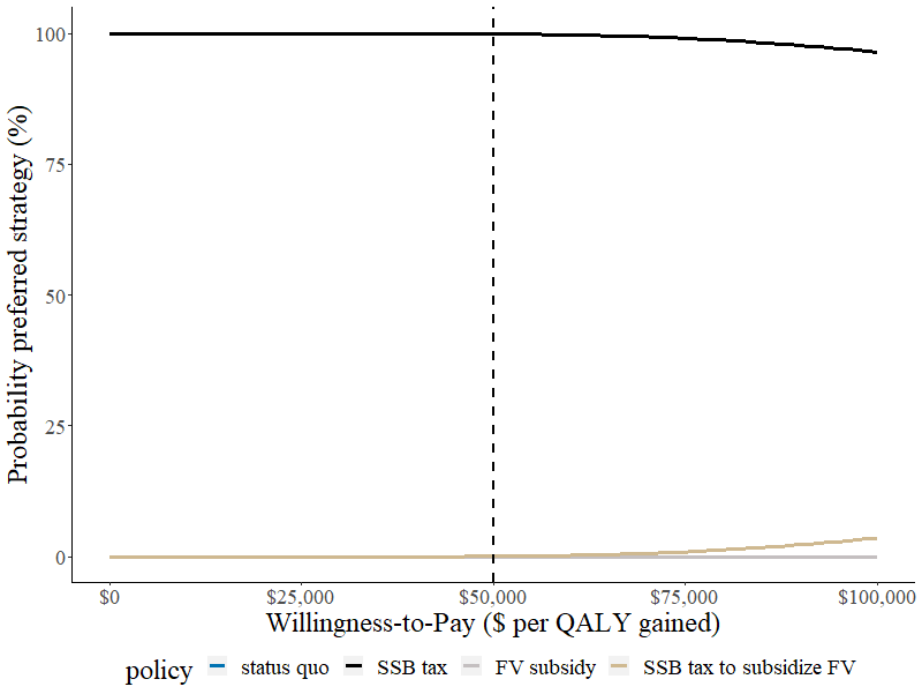


Figure B.4a. Cost-Effectiveness Acceptance Curve Plot Over 10 Years from Healthcare Perspective

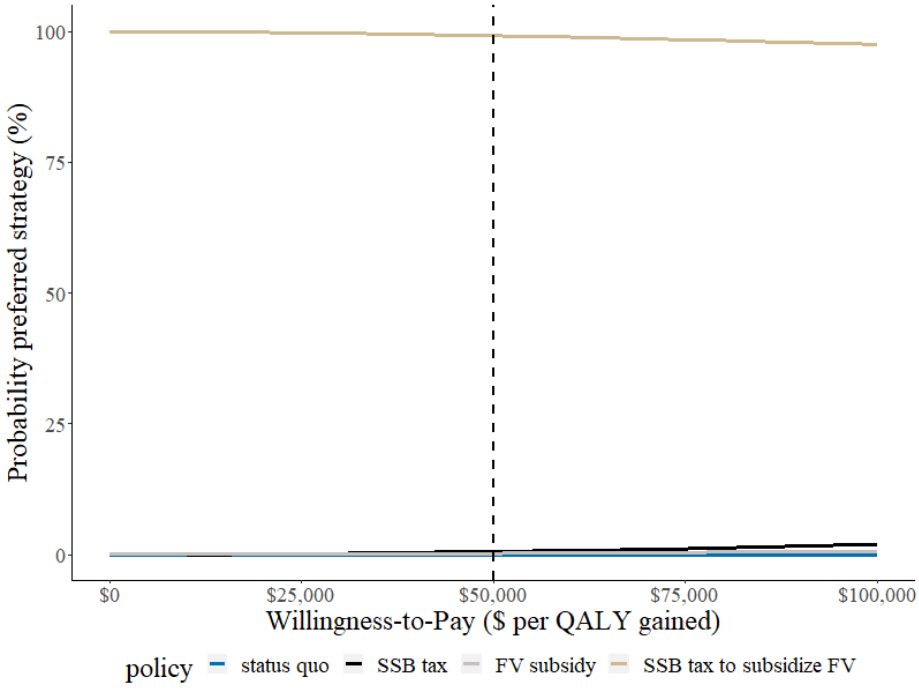


Figure B.4b. Cost-Effectiveness Acceptance Curve Plot Over 10 Years from Societal Perspective

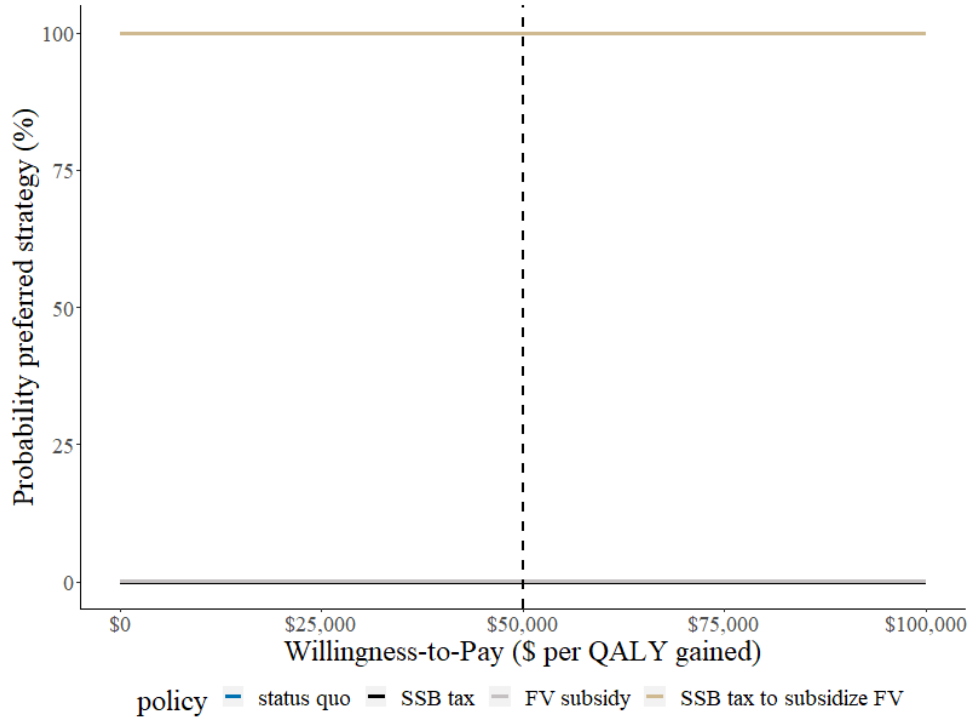


Figure B.4c. Cost-Effectiveness Acceptance Curve Plot Over Lifetime from Healthcare Perspective

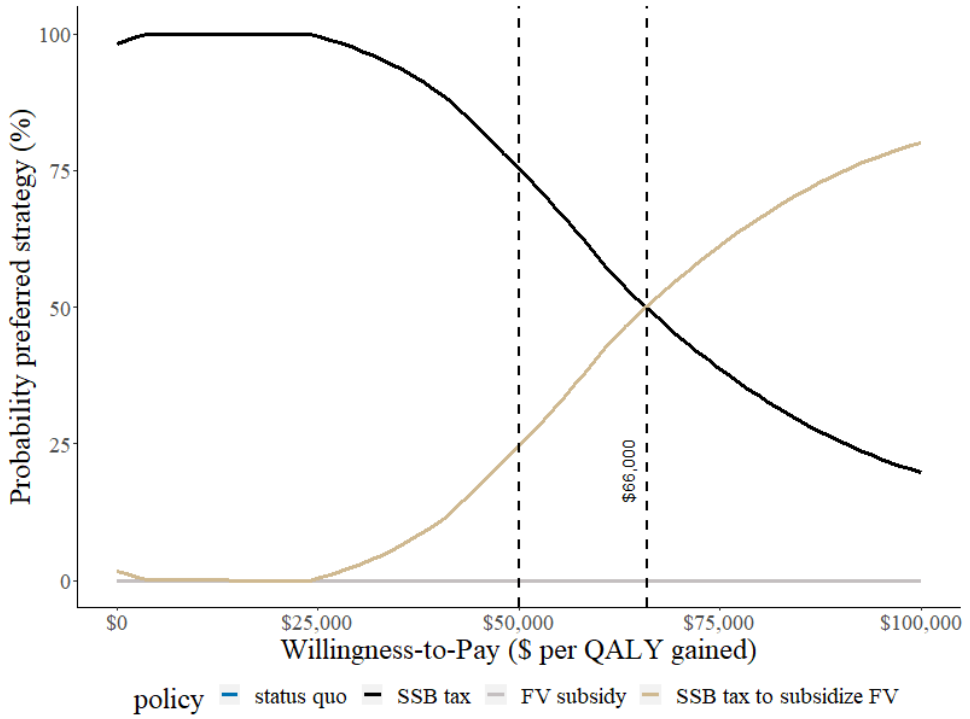


Figure B.4d. Cost-Effectiveness Acceptance Curve Plot Over Lifetime from Societal Perspective

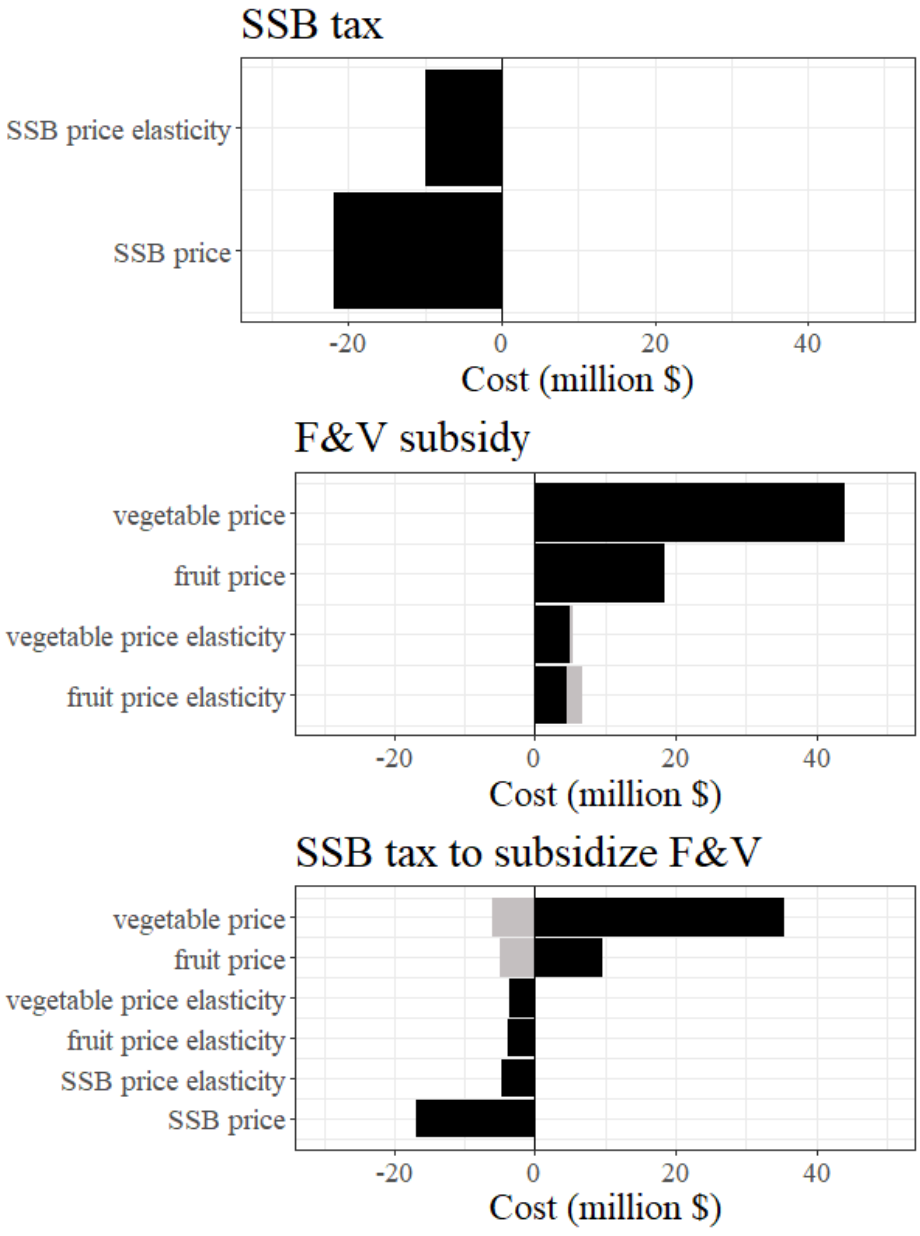


Figure B.5. One-Way Sensitivity Analysis Tornado Diagram on Food Prices and Food Price Elasticities.

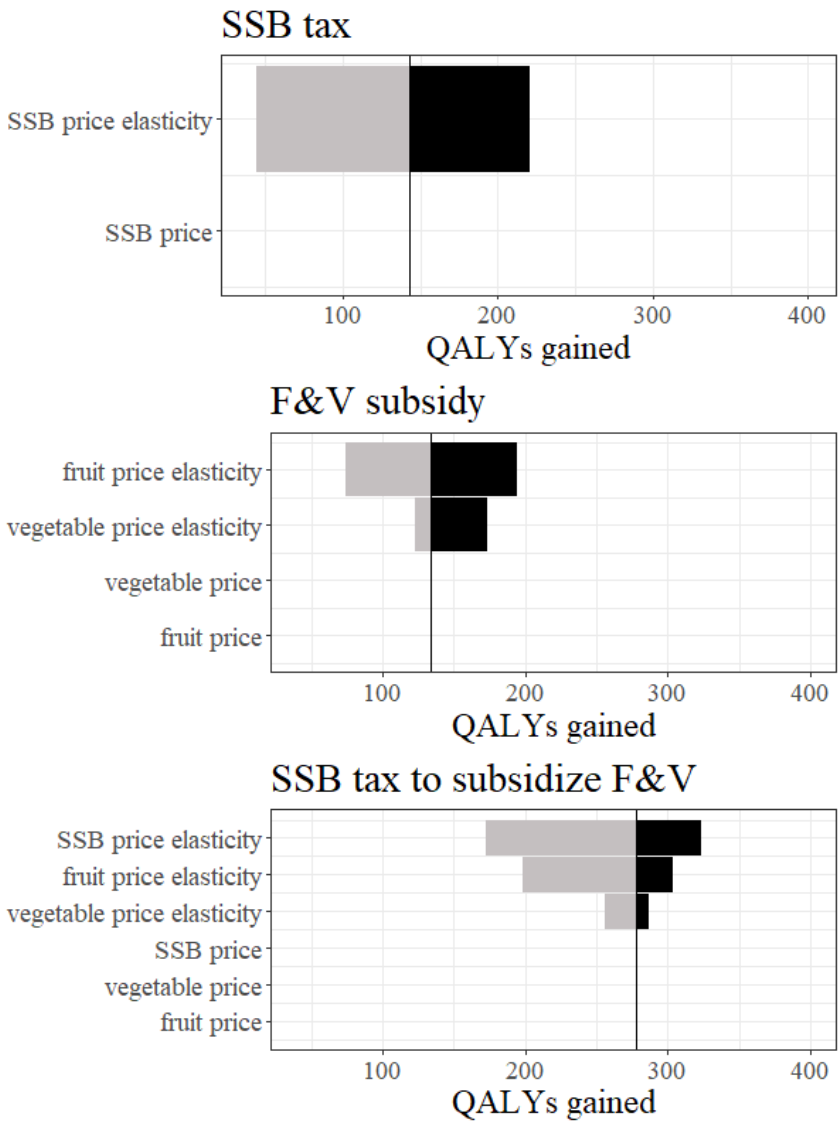


Figure B.5. continued.

Vegetable price elasticity (0.36, 0.58, 0.82) – (low value, base value, high value)
Vegetable price $ per serving (0.07, 0.32, 2.18)
Fruit price elasticity (0.22, 0.7, 1.09)
Fruit price $ per serving (0.15, .26, 1.3)
SSB price elasticity (0.2, 0.79, 1.3)
SSB price $ per serving (0.27, 0.33, 1.21)
The tornado diagrams show the impact that independently changing the model parameters has on the costs and QALYs for each individual treatment vs. status quo.

Model Validation Results Against CVD Policy Model

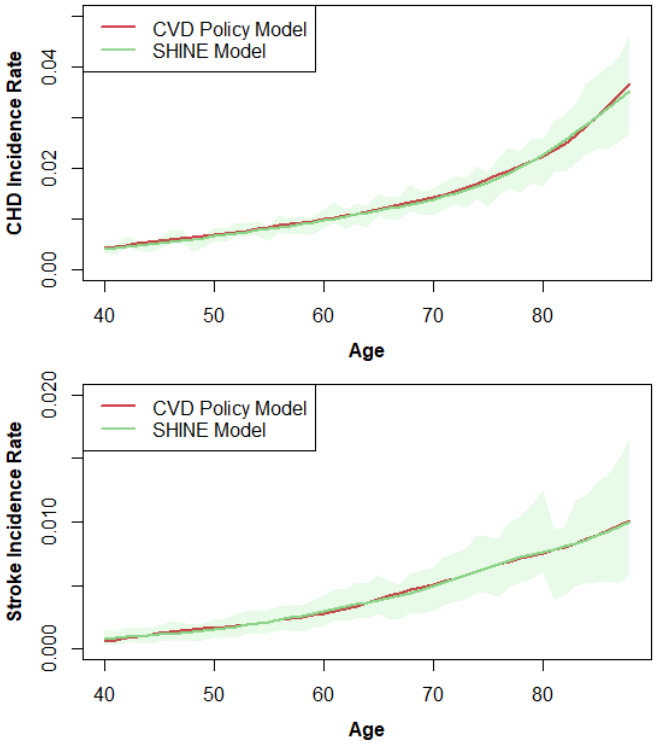


Figure B.6. CHD and Stroke Incidence Rate by Male

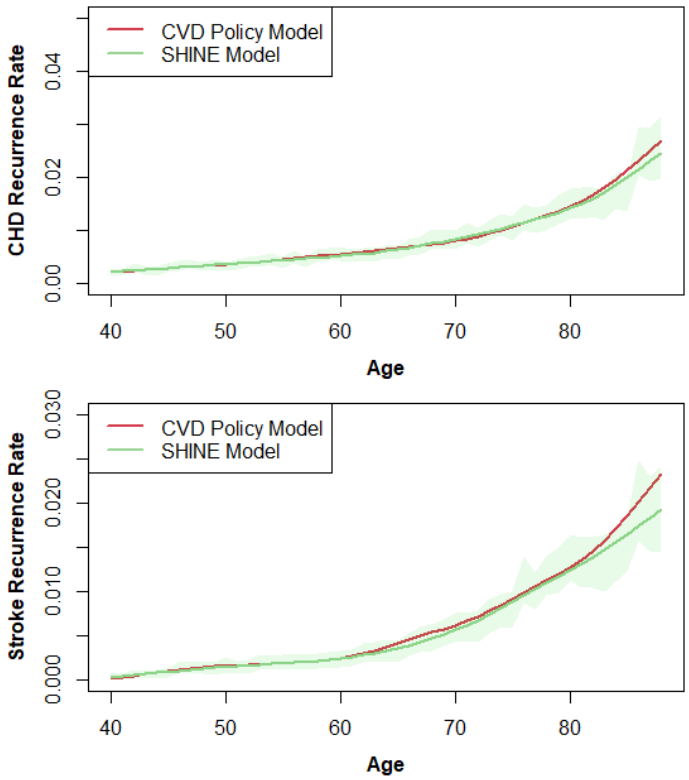


Figure B.7. CHD and Stroke Incidence Rate by Female

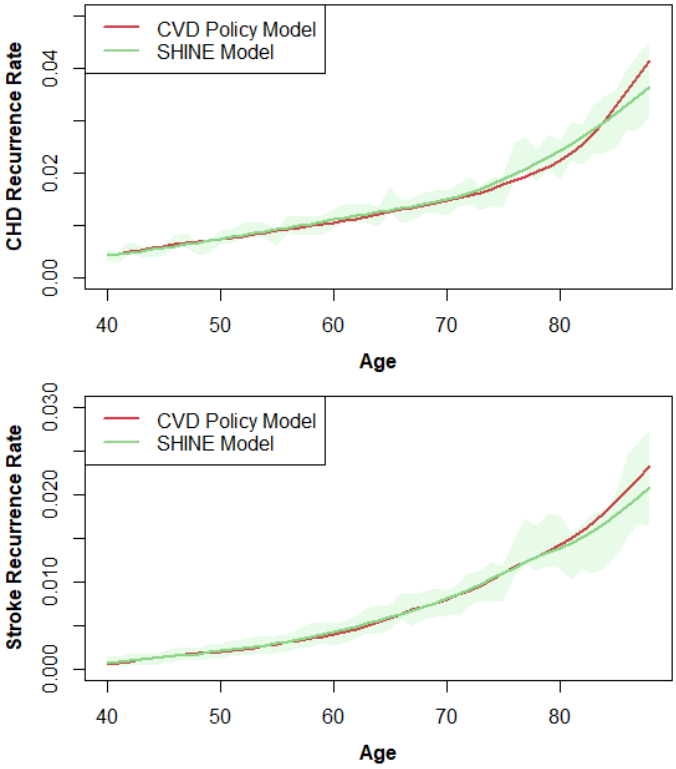


Figure B.8. CHD and Stroke Recurrence Rate by Male

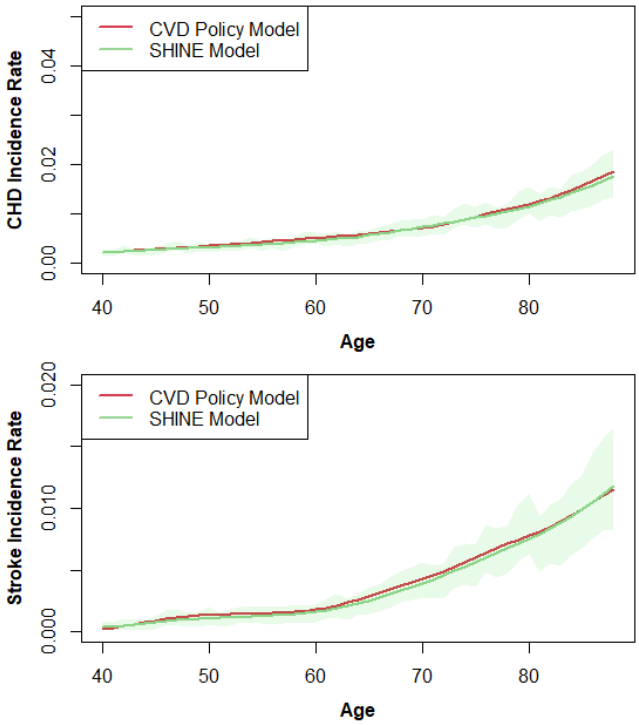


Figure B.9. CHD and Stroke Recurrence Rate by Female

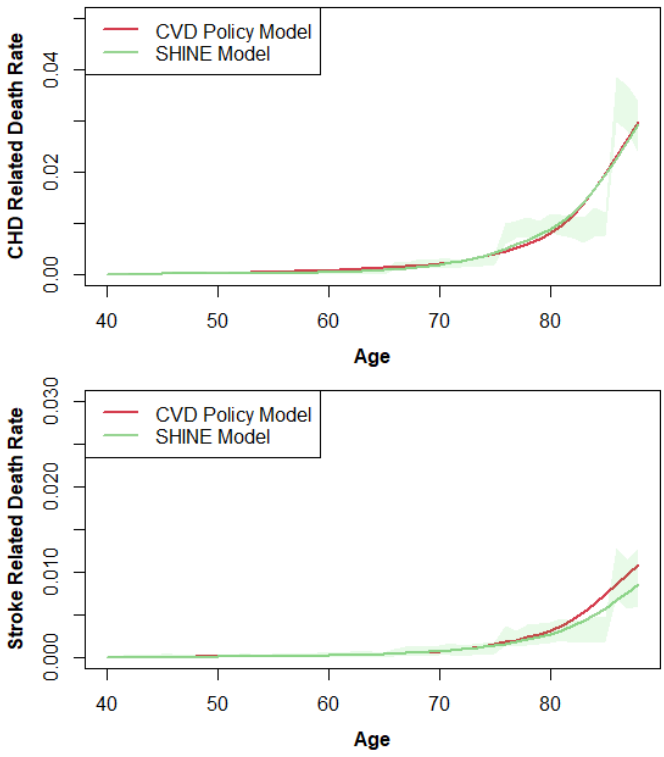


Figure B.10. CHD and Stroke Related Death Rate by Male

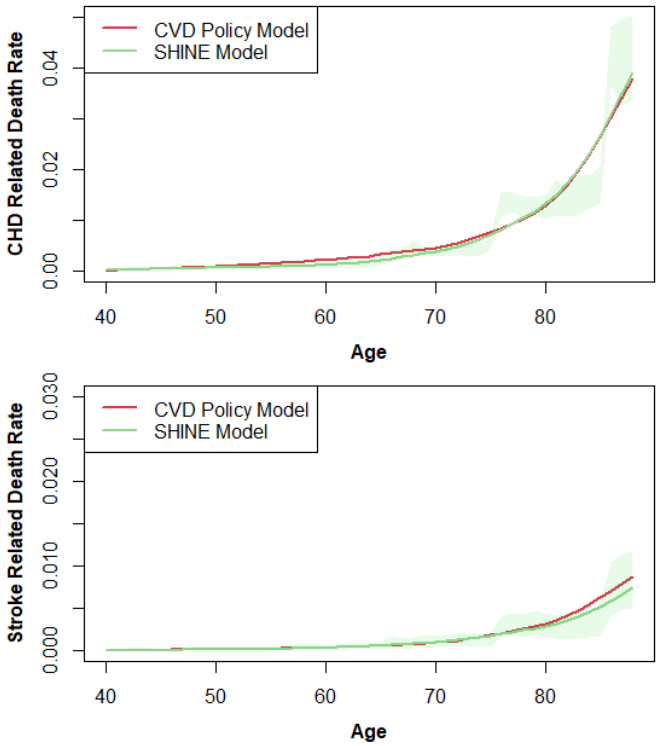


Figure B.11. CHD and Stroke Related Death Rate by Female

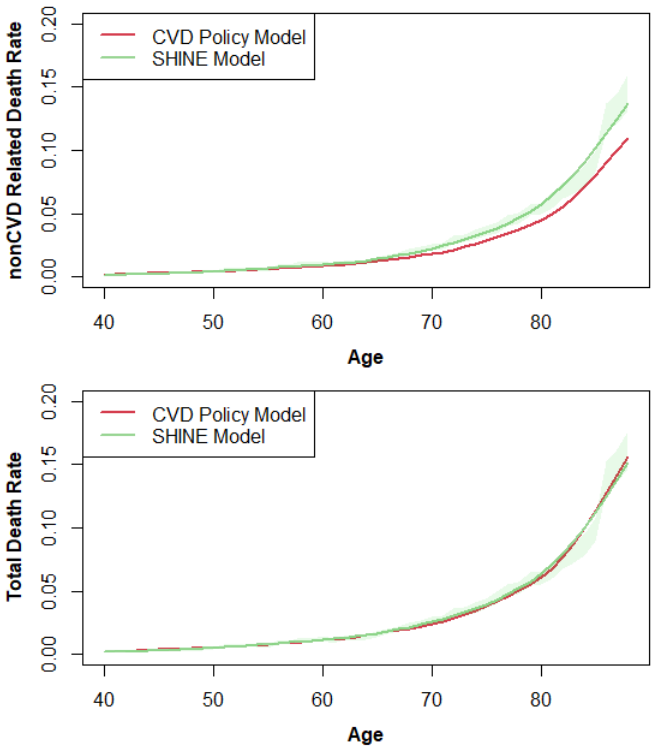


Figure B.12. Non-CVD Related Death and Total Death Rate by Female

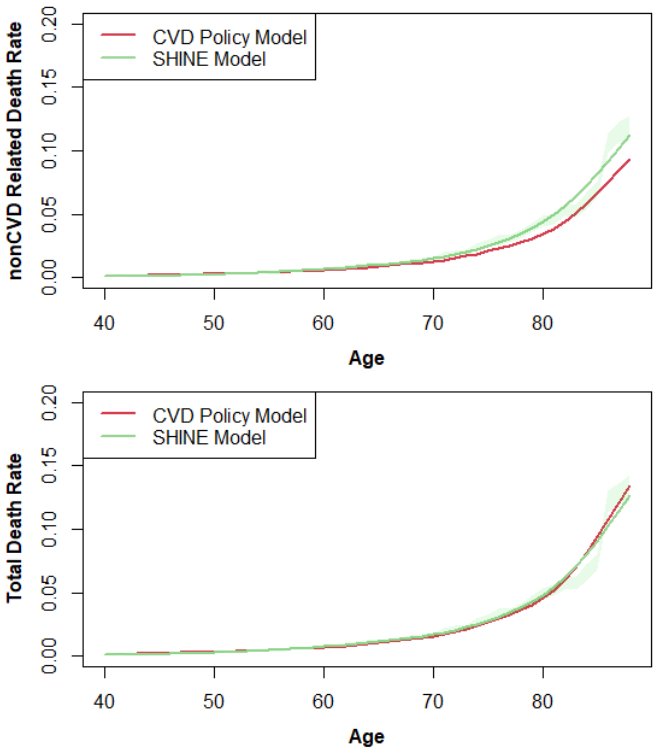


Figure B.13. Non-CVD Related Death and Total Death Rate by Male

Table B.1. Logistic **Risk** Functions used to Determine Parameter Coefficient for Probability of Incident Event

$$P(Incident\ Event) = \frac{\exp\left(\alpha + \beta_{p1}P1 + \beta_{P2}P2 + \cdots + \beta_{Pn}Pn\right)}{1 + \exp\left(\alpha + \beta_{p1}P1 + \beta_{P2}P2 + \cdots + \beta_{Pn}Pn\right)}$$

Source: Columbia University-National Heart Lung and Blood Institute Pooled Cohorts Dataset[89]

Parameters	Description	Hazard Ratio	Beta Values
Parameters for Probability of Incident Coronary Heart Disease Event			
Age	Years (Age – 55) (centered around 55)	1.107	0.1016
Black	Binary	0.885	-0.1219
BMI	kg/m^2	1.006 (1.000, 1.012)	0.00597
Former Smoker	Binary	1.204 (1.134, 1.278)	0.1857
Current Smoker	Binary	1.683 (1.496, 1.893)	0.5205
Cigarettes per day	Among current smokers	1.006 (1.001, 1.011)	0.006
Systolic Blood Pressure (SBP)	mmHg	1.013 (1.012, 1.014)	0.0129
Diabetes	Binary	1.916 (1.789, 2.052)	0.6503
HDL-c	mg/dL	0.985 (0.983, 0.988)	-0.0149
LDL-c	mg/dL	1.005 (1.005, 1.006)	0.0054
eGFR	$mL/min/1.73^2$	0.993 (0.992, 0.995)	-0.0068
Age x Current Smoker	-	0.987 (0.982, 0.991)	-0.0135
Age x SBP	-	1.000 (1.000, 1.000)	-0.0003
Age x Diabetes	-	0.990 (0.985, 0.995)	-0.0103
Age x HDL-c	-	1.000 (1.000, 1.000)	0.0003
Age x LDL-c	-	1.000	-0.0002
Parameters for Probability of Incident Stroke Event			
Age	Years (Age – 55) (centered around 55)	1.146 (1.123, 1.170)	0.1366
Black	Binary	1.605 (1.430, 1.802)	0.4733
Current Smoker	Binary	1.868 (1.667, 2.094)	0.6251
Systolic Blood Pressure (SBP)	mmHg	1.020 (1.018, 1.022)	0.0199
Diabetes	Binary	1.950 (1.751, 2.171)	0.6677
HDL-c	mg/dL	0.995 (0.992, 0.998)	-0.0047
LDL-c	mg/dL	1.002 (1.000, 1.003)	0.0017
eGFR	$mL/min/1.73^2$	0.996 (0.993, 0.998)	-0.0042
Age x Black	-	0.977 (0.969, 0.986)	-0.0228

Parameters	Description	Hazard Ratio	Beta Values
Age x Current Smoker	-	0.990 (0.982, 0.999)	-0.0096
Age x SBP	-	1.000 (0.999, 1.000)	-0.0004
Age x Diabetes	-	0.984 (0.977, 0.991)	-0.0161
Parameters for Probability of Non-Cardiovascular Disease Mortality			
Age	Years (Age – 55) (centered around 55)	1.104 (1.097, 1.111)	0.0992
Black	Binary	1.501 (1.404, 1.605)	0.4064
BMI	kg/m^2	0.905 (0.886, 0.925)	-0.0996
BMI^2	$(kg/m^2)^2$	1.001 (1.001, 1.002)	0.0014
Former Smoker	Binary	1.296 (1.228, 1.369)	0.2597
Current Smoker	Binary	1.985 (1.792, 2.200)	0.6859
Cigarettes per day	Among current smokers	1.020 (1.016, 1.025)	0.0203
Systolic Blood Pressure (SBP)	mmHg	1.001 (1.000, 1.002)	0.0011
Diabetes	Binary	1.542 (1.441, 1.650)	0.433
eGFR	$mL/min/1.73^2$	0.993 (0.992, 0.995)	-0.0066
Age x Black	-	0.985 (0.980, 0.989)	-0.0153
Age x BMI^2	-	1.000 (1.000, 1.000)	0.00002
Age x Diabetes	-	0.989 (0.984, 0.994)	-0.0114

BMI – body mass index, UI – confidence interval, eGFR – Estimated Glomerular Filtration Rate, HDL-C – high-density

Table B.2. Probabilities for non-incident cardiovascular disease events.

Parameter	Base Case Value (%)	Source(s)
Following CHD event (annual probability)		
Recurrent[a] CHD event within 1 year of previous CHD event		
Men		
40-44 years	3.53	
45-54 years	4.74	
55-64 years	6.49	
65-74 years	7.96	
75+ years	12.8	150–153
Women		
40-44 years	2.26	
45-54 years	3.96	
55-64 years	4.98	
65-74 years	8.29	
75+ years	13.55	
Recurrent CHD event after 1 year of previous CHD event		
Men		
40-44 years	1.22	
45-54 years	1.60	
55-64 years	2.23	
65-74 years	2.79	
75+ years	4.53	150–154
Women		
40-44 years	0.96	
45-54 years	1.25	
55-64 years	1.63	
65-74 years	2.72	
75+ years	4.66	
Stroke after CHD		
Men		
40-44 years	0.55	
45-54 years	0.55	
55-64 years	0.79	
65-74 years	0.83	
75+ years	0.92	155,156
Women		
40-44 years	0.55	
45-54 years	0.55	
55-64 years	0.77	
65-74 years	0.87	
75+ years	0.89	
Following stroke event (annual probability)		

Parameter	Base Case Value (%)	Source(s)
Recurrent stroke event	3.6	[157]
CHD after stroke within 10 years	2.5	[158]
CHD after stroke after 10 years	2.2	[159]
30-day case fatality rates		
Incident CHD		
Men		[153,160–163]
40-44 years	9.37	
45-54 years	14.6	
55-64 years	17.44	
65-74 years	20.77	
75-85 years	18.41	
85+ years	78.18	
Women		
40-44 years	7.08	
45-54 years	9.83	
55-64 years	13.16	
65-74 years	17.97	
75-85 years	14.97	
85+ years	81.39	
Recurrent CHD		
Men		[153,160–163]
40-44 years	2.24	
45-54 years	7.84	
55-64 years	9.89	
65-74 years	12.96	
75-85 years	14.6	
85+ years	27.16	
Women		
40-44 years	2.22	
45-54 years	5.44	
55-64 years	6.65	
65-74 years	11.48	
75-85 years	10.95	
85+ years	75.79	
Any stroke		
Men		[164]
40-44 years	6.23	
45-54 years	7.55	
55-64 years	8.95	
65-74 years	13.88	
75-85 years	21.2	
85+ years	37.5	

Parameter	Base Case Value (%)	Source(s)
Women		
40-44 years	13.7	
45-54 years	7.45	
55-64 years	10.65	
65-74 years	11.92	
75-85 years	23.02	
85+ years	46.5	
Other		
Maximum annual number of CVD events per cycle	2	Assumption

[a]Recurrent event occurs subsequent to primary CHD or stroke event
CHD – coronary heart disease, CVD – cardiovascular disease

Table B.3. Population Characteristics of NYC HANES pooled population

Parameters	NYCHANES (95% UI*)	NYCHANES Matched with Pooled Cohort (95% UI)	% of Sample mean within 95%UI	Sampled Population (95% UI)
Male	46.4% (45.3% , 47.6%)	37.1% (35.2% , 38.9%)	100	37.1% (37% , 37.1%)
Age	45.8 (44.7, 46.9)	50 (47.5, 52.5)	100	50 (50, 50)
African American	21.2% (27% , 21.7%)	24.3% (23.1% , 25.5%)	99.9	24.3% (24.3% , 24.3%)
BMI	28 (27.3 , 28.7)	26.5 (25.1, 27.8)	100	26.5 (26.5, 26.5)
Current Smoker	19.1% (18.6% , 19.6%)	24.2% (23% , 25%)	99.4	24.2% (24.2% , 24.2%)
Diastolic blood pressure	73.7 (71.9 , 75.6)	71.1 (67.6, 74.7)	100	71.1 (71.1 , 71.1)
Systolic blood pressure	120.1 (117.1 , 123.1)	121 (115, 127.1)	100	121 (121 , 121)
DM	10.6% (10.3% , 10.9%)	18% (17%, 19%)	97.8	18% (18% , 18.1%)
HDLC	56.6 (55.2, 58)	50.9 (48.3, 53.4)	100	50.9 (50.9, 50.9)
LDLC	109.8 (107, 112.5)	111.7 (106.1, 117.3)	100	111.7 (111.7 , 111.7)

*UI: uncertainty interval, equivalent to confidence interval.

Table B.4. Fruit, Vegetable and SSB Consumption

Gender	Race	Fruits		Vegetables		SSB		Source
		Mean	*sd*	*Mean*	*sd*	*Mean*	*sd*	
Male	African American	0.93	1.06	1.04	0.99	1.32	2.81	Analysis of Community Health Survey (CHS) [165]
	Others	1.18	1.28	1.34	1.31	0.72	1.56	
Female	African American	1.01	1.10	1.25	1.23	0.99	1.78	
	Others	1.25	1.09	1.49	1.53	0.38	0.94	

Consumption (servings per day) (Truncated Normal Distribution*)

Table B.5. Food prices and price elasticities by Fruit, Vegetable and SSB

Parameter	Distribution	Mean (Sampled Range)	Source
Fruit price	Gamma Distribution	0.28 (0.15, 1.3)	[166]
Vegetable Price		0.32 (0.07, 2.18)	
SSB Price		0.33 (0.27, 1.21)	[167]
Fruit Price Elasticity	Truncated Normal Distribution	0.7 (0.22, 1.09)	[147]
Vegetable Price Elasticity		0.58 (0.36, 0.82)	
SSB Price Elasticity		0.79 (0.21, 1.29)	[168]

Table B.6. Reduced Risk due to Consumption Change by Fruit, Vegetable and SSB

$$RR = \frac{RR}{serving/day}^{New\ Intake - Original\ Intake}$$

Event	Distribution	Mean RR for per serving reduction / day (95% UI)	Source
Fruit			
CHD	Log-normal	0.93 (0.89, 0.97)	[169]
Stroke		0.82 (0.75, 0.91)	[170]
DM		0.94 (0.89, 1.00)	
Vegetable			
CHD	Log-normal	0.95 (0.92, 0.98)	[169]
Stroke		0.94 (0.90, 0.99)	[170]
DM		0.98 (0.89, 1.08)	
SSB			
CHD	Log-normal	0.93 (0.89, 0.97)	[169]
Stroke		0.82 (0.75, 0.91)	[171]
DM		0.22 (0.08, 0.36)	[169]

Table B.7**.** Cost of Health State and Acute Event

Cost (2019 USD)	Value	Source
Background health cost		
Men		
40-49 years	3,689	
50-59 years	4,849	
60-69 years	6,461	
70-79 years	9,609	
80-89 years	14,541	
90+ years	27,874	[172]
Women		
40-49 years	5,183	
50-59 years	7,034	
60-69 years	10,120	
70-79 years	12,426	
80-89 years	18,528	
90+ years	32,515	
CHD first year		
Aged 40-69	13,273	[172]
Aged 70+	20,284	
CHD subsequent years		
Aged 40-89	2,711	[172]
Aged 90+	4,262	
Acute (30-day) CHD		
Men		
40-49 years	8,317	
50-59 years	14,135	
60-69 years	20,454	
70-79 years	24,131	
80-89 years	25,174	
90+ years	26,258	[173,174]
Women		
40-49 years	6,608	
50-59 years	8,874	
60-69 years	17,312	
70-79 years	22,112	
80-89 years	25,957	
90+ years	34,502	
CHD Mortality		
Men		
40-49 years	64,209	
50-59 years	67,520	[173,174]
60-69 years	73,412	
70-79 years	64,513	

Cost (2019 USD)	Value	Source
80-89 years	54,473	
90+ years	46,475	
Women		
40-49 years	64,614	
50-59 years	56,959	
60-69 years	69,176	
70-79 years	63,939	
80-89 years	54,640	
90+ years	46,274	
Stroke first year		
All ages	20,538	[172]
Stroke subsequent years		
All ages	5,707	[172]
Acute (30-day) stroke		
Men		
40-49 years	26,171	
50-59 years	22,736	
60-69 years	21,228	
70-79 years	17,915	
80+ years	19,144	[173,174]
Women		
40-49 years	25,278	
50-59 years	21,842	
60-69 years	20,336	
70-79 years	17,023	
80+ years	18,251	
Stroke Mortality		
Men		
40-49 years	32,344	
50-59 years	30,070	
60-69 years	28,724	
70-79 years	25,763	
80+ years	26,861	[173,174]
Women		
40-49 years	32,344	
50-59 years	29,272	
60-69 years	27,926	
70-79 years	24,965	
80+ years	26,063	
Inflation factor		
$US2010 to $US2019	1.2587	[175]

CHD – coronary heart disease

Table B.8. Chronic and Acute Utilities

Parameter	Value	Source
CHD		
Age 40-44	0.9348	176–178
Age 45-54	0.9374	
Age 55-64	0.9376	
Age 65-74	0.9372	
Age 75-84	0.9364	
Age 85+	0.9358	
Stroke		
All ages	0.8835	176–178
Acute (30-day) CHD		
Age 40-44	0.897	176–178
Age 45-54	0.8862	
Age 55-64	0.8669	
Age 65-74	0.8351	
Age 75-84	0.7946	
Age 85+	0.6829	
Acute (30-day) stroke		
All ages	0.8662	176–178

CHD – coronary heart disease, CVD – cardiovascular disease events

Table B.9. Reporting Checklist for Economic Evaluation of Health Interventions.

Based on the CHEERS 2022 Checklist, available from https://www.equator-network.org/wp-content/uploads/2013/04/CHEERS-2022-checklist-1.pdf

	Item	Guidance for Reporting	Reported in section
Title			
Title	#1	Identify the study as an economic evaluation and specify the interventions being compared.	Title
Abstract			
Abstract	#2	Provide a structured summary that highlights context, key methods, results and alternative analyses.	Abstract
Introduction			
Background and objectives	#3	Give the context of the study, the study question and its practical relevance for decision making in policy or practice.	Introduction
Methods			
Health economic analysis plan	#4	Indicate whether a health economic analysis plan was developed and where available.	No
Study population	#5	Describe characteristics of the study population (such as age range, demographics, socioeconomic, or clinical characteristics).	Simulated Population in Methods
Setting and location	#6	Provide relevant contextual information that may influence findings	Methods
Comparators	#7	Describe the interventions or strategies being compared and why chosen.	Policy Scenarios in Methods
Perspective	#8	State the perspective(s) adopted by the study and why chosen	Statistical Analysis in Methods
Time horizon	#9	State the time horizon for the study and why appropriate.	Model Development in Methods
Discount rate	#10	Report the discount rate and reason chosen	Statistical Analysis in Methods
Selection of outcomes	#11	Describe what outcomes were used as the measure(s) of benefit and harm(s).	Cost and Utility Model Parameters in Methods
Measurement of outcomes	#12	Describe how outcomes used to capture benefit(s) and harm(s) were measured.	Cost and Utility Model Parameters in Methods

Valuation of outcomes	#13	Describe the population and methods used to measure and value outcomes.	Cost and Utility Model Parameters in Methods
Measurement and valuation of resources and costs	#14	Describe how costs were valued.	Cost and Utility Model Parameters in Methods and Model Inputs in Appendix B
Currency, price date, and conversion	#15	Report the dates of the estimated resource quantities and unit costs, plus the currency and year of conversion	Cost and Utility Model Parameters in Methods
Rationale and description of model	#16	If modelling is used, describe in detail and why used. Report if the model is publicly available and where it can be accessed.	Model Development in Methods and Model Overview in Appendix B
Analytics and Assumptions	#17	Describe any methods for analyzing or statistically transforming data, any extrapolation methods, and approaches for validating any model used.	Model Overview in Appendix B
Characterizing heterogeneity	#18	Describe any methods used for estimating how the results of the study vary for sub-groups.	Simulated Population in Methods and Model Overview in Appendix B
Characterizing distributional effects	#19	Describe how impacts are distributed across different individuals or adjustments made to reflect priority populations.	Table S11 in Appendix B
Characterizing uncertainty	#20	Describe methods to characterize any sources of uncertainty in the analysis.	Statistical Analysis in Methods
Approach to engagement with patients and others affected by the study	#21	Describe any approaches to engage patients or service recipients, the general public, communities, or stakeholders (e.g., clinicians or payers) in the design of the study.	No
Results			
Study parameters	#22	Report all analytic inputs (e.g., values, ranges, references) including uncertainty or distributional assumptions.	Table S1-S8, S12 in Appendix B
Summary of main results	#23	Report the mean values for the main categories of costs and outcomes of interest and summarize them in the most appropriate overall measure.	Results, Table 3.1

Effect of uncertainty	#24	Describe how uncertainty about analytic judgments, inputs, or projections affect findings. Report the effect of choice of discount rate and time horizon, if applicable.	Results, Figure 3.2-3
Effect of engagement with patients and others affected by the study	#25	Report on any difference patient/service recipient, general public, community, or stakeholder involvement made to the approach or findings of the study	No
Discussion			
Study findings, limitations, generalizability, and current knowledge	#26	Report key findings, limitations, ethical or equity considerations not captured, and how these could impact patients, policy, or practice.	Discussion
Other			
Source of funding	#27	Describe how the study was funded and any role of the funder in the identification, design, conduct, and reporting of the analysis.	Funding
Conflict of interest	#28	Report authors conflicts of interest according to journal or International Committee of Medical Journal Editors requirements.	Cover Letter

Table B.10. Health Outcomes and Cost-Effectiveness over Multiple Time Horizon

	SSB Tax only	F&V Subsidy only	SSB tax to subsidize F&V
Over 20 years			
Prevented Healthcare Outcomes			
CHD events	34 (10 to 63)	16 (0 to 33)	49 (20 to 82)
Stroke events	17 (4 to 32)	15 (4 to 29)	32 (15 to 52)
CVD deaths	8 (-2 to 21)	4 (-4 to 14)	14 (0 to 29)
QALYs gained	64 (12 to 129)	39 (4 to 82)	101 (35 to 177)
Incremental Costs (2019 USD, thousands)			
Healthcare sector perspective	-1,540 (-2,730 to -610)	-980 (-1,590 to -470)	-2,490 (-3,690 to -1,430)
Societal perspective	-6,930 (-7,870 to -6,190)	5,470 (4,880 to 5,970)	-1,430 (-2,500 to 580)
Incremental cost-effectiveness			
Healthcare sector perspective			
ICER	Dominated*	Dominated*	Dominant**
INMB	4,613,291 (1,474,329 to 8,600,976)	2,881,597 (992,333 to 5,468,697)	7,403,528 (3,534,993 to 12,258,909)
Societal perspective			
ICER	Dominant**	Dominated***	Dominant**
INMB	10,007,844 (7,097,779 to 13,839,082)	-3,576,846 (-5,433,216 to -1,010,781)	6,338,897 (2,663,346 to 10,974,245)
Over 40 years			
Prevented Healthcare Outcomes			
CHD events	61 (20 to 107)	30 (8 to 56)	91 (44 to 143)
Stroke events	35 (12 to 62)	32 (13 to 55)	66 (36 to 101)
CVD deaths	19 (2 to 40)	10 (-3 to 25)	31(7 to 56)
QALYs gained	142 (44 to 267)	88 (23 to 171)	227 (102 to 369)
Incremental Costs (2019 USD, thousands)			
Healthcare sector perspective	-2,730 (-4,550 to -1,140)	-1,790 (-2,740 to -960)	-4,480 (-6,460 to -2,640)
Societal perspective	-10,140 (-11,630 to -8,910)	7,120 (6,210 to 7,920)	-3,000 (-4,750 to -1,590)
Incremental cost-effectiveness			
Healthcare sector perspective			
ICER	Dominated*	Dominated*	Dominant**
INMB	9,788,639 (3,494,787 to 17,204,221)	6,213,658 (2,707,155 to 10,541,790)	15,780,529 (8,384,293 to 24,183,271)
Societal perspective			

	SSB Tax only	F&V Subsidy only	SSB tax to subsidize F&V
ICER	Dominant**	Dominated***	Dominant**
INMB	17,207,776 (11,279,099 to 24,352,009)	-2,694,328 (-6,179,486 to 1,636,413)	14,289,757 (7,230,732 to 22,452,060)
Over lifetime			
Prevented Healthcare Outcomes			
CHD events	79 (28 to 139)	39 (11 to 71)	118 (60 to 187)
Stroke events	49 (17 to 89)	46 (19 to 80)	94 (54 to 145)
CVD deaths	41 (8 to 81)	25 (1 to 51)	67 (24 to 114)
QALYs gained	196 (67 to 351)	122 (44 to 224)	315 (159 to 495)
Incremental Costs (2019 USD, thousands)			
Healthcare sector perspective	-3,230 (-5,480 to -1,400)	-2,160 (-3,330 to -1,150)	-5,340 (-7,690 to -3,210)
Societal perspective	-11,220 (-13,090 to -9,790)	7,440 (6,340 to 8,420)	-3,730 (-5,770 to -2,060)
Incremental cost-effectiveness			
Healthcare sector perspective			
ICER	Dominated*	Dominated*	Dominant**
INMB	13,030,000 (8,830,000 to 18,950,000)	8,260,000 (5,530,000 to 12,350,000)	21,090,000 (15,640,000 to 27,960,000)
Societal perspective			
ICER	Dominant**	Dominated***	Dominant**
INMB	21,020,000 (16,440,000 to 27,340,000)	-1,340,000 (-4,140,000 to 2,780,000)	19,480,000 (13,720,000 to 26,810,000)

Notes: All numerical results are presented as mean estimate with (95% UIs)

*Dominated (i.e., costs more and less effective) by combining SSB taxation with F&V subsidies

**Dominant (i.e., costs less and more effective) vs. status quo

***Dominated by SSB taxation alone and combining SSB taxation with F&V subsidies

Table B.11. Health Outcomes and Cost-Effectiveness by Subpopulation Over 10 Years

Each subgroup population size is 10K.

	SSB Tax only	F&V Subsidy only	SSB tax to subsidize F&V
Male			
Prevented Healthcare Outcomes			
CHD events	28 (4 to 59)	11 (-4 to 29)	39 (9 to 73)
Stroke events	11 (0 to 26)	8 (0 to 20)	18 (4 to 36)
CVD deaths	8 (-4 to 27)	4 (-6 to 16)	11 (-4 to 34)
QALYs gained	16 (0 to 38)	8 (-3 to 25)	23 (2 to 48)
Incremental Costs (2019 USD, thousands)			
Healthcare sector perspective	-1,140 (-2,220 to -320)	-540 (-1,080 to -120)	-1,680 (-2,800 to -120)
Societal perspective	-5,310 (-6,760 to -4,120)	3,250 (2,580 to 3,790)	-2,060 (-3,570 to -150)
Incremental cost-effectiveness			
Healthcare sector perspective			
ICER	Dominated*	Dominated*	Dominant**
INMB	2,938,327 (537,267 to 6,208,130)	1,396,394 (67,126 to 3,489,340)	4,247,183 (1,287,951 to 7,911,067)
Societal perspective			
ICER	Dominant**	Dominated***	Dominant**
INMB	7,109,092 (4,816,401 to 10,312,578)	-2,390,236 (-3,700,959 to -282,338)	4,630,950 (1,756,887 to 8,193,287)
Female			
Prevented Healthcare Outcomes			
CHD events	9 (-2 to 25)	7 (-5 to 20)	16 (0 to 36)
Stroke events	6 (0 to 16)	8 (0 to 18)	13 (4 to 27)
CVD deaths	2 (-4 to 10)	3 (-6 to 10)	5 (-6 to 16)
QALYs gained	8 (-2 to 24)	7 (-2 to 20)	14 (0 to 34)
Incremental Costs (2019 USD, thousands)			
Healthcare sector perspective	-430 (-910 to -80)	-410 (-830 to -80)	-810 (-1,460 to -330)
Societal perspective	-2,990 (-3,690 to-2,430)	3,660 (3,130 to 4,120)	700 (-190 to 1,420)
Incremental cost-effectiveness			
Healthcare sector perspective			
ICER	Dominated*	Dominated*	Dominant**
INMB	1,116,815 (115,749 to 2,832,847)	1,059,440 (108,962 to 2,600,640)	2,106,028 (515,048 to 4,416,239)

	SSB Tax only	F&V Subsidy only	SSB tax to subsidize F&V
Societal perspective			
ICER	Dominant**	Dominated***	$50,000
INMB	3,680,361 (2,720,549 to 5,337,711)	-3,015,641 (-4,004,996 to -1,482,978)	594,361 (-969,168 to 2,769,523)
White			
Prevented Healthcare Outcomes			
CHD events	17 (0 to 42)	9 (-8 to 30)	27 (0 to 57)
Stroke events	7 (0 to 19)	8 (0 to 19)	14 (0 to 31)
CVD deaths	3 (-6 to 14)	2 (-5 to 14)	5 (-8 to 19)
QALYs gained	8 (-3 to 24)	6 (-4 to 19)	13 (-2 to 31)
Incremental Costs (2019 USD, thousands)			
Healthcare sector perspective	-670 (-1,450 to -110)	-480 (-1,060 to -50)	-1,150 (-2,060 to -390)
Societal perspective	-3,550 (-4,500 to -2,620)	3,540 (2,810 to 4,100)	90 (-1,090 to 1,120)
Incremental cost-effectiveness			
Healthcare sector perspective			
ICER	Dominated*	Dominated*	Dominant**
INMB	1,707,336 (-66,305 to 4,465,762)	1,246,979 (-58,868 to 3,472,100)	2,868,731 (241,152 to 6,118,718)
Societal perspective			
ICER	Dominant**	Dominated***	$7,000
INMB	4,486,826 (2,728,935 to 7,190,682)	-2,772,658 (-4,089,538 to -471,930)	1,628,341 (-886,785 to 4,826,790)
Black			
Prevented Healthcare Outcomes			
CHD events	27 (-5 to 68)	7 (-14 to 29)	34 (-5 to 76)
Stroke events	15 (0 to 42)	8 (0 to 24)	23 (0 to 48)
CVD deaths	5 (-10 to 28)	2 (-10 to 18)	7 (-14 to 33)
QALYs gained	8 (-3 to 25)	3 (-5 to 13)	11 (-3 to 30)
Incremental Costs (2019 USD, thousands)			
Healthcare sector perspective	-1,240 (-2,680 to -210)	-440 (-1,190 to 10)	-1,660 (-3,240 to -500)
Societal perspective	-6,360 (-8,340 to -4,790)	2,860 (1,950 to 3,470)	-3,470(-5,590 to -1,780)
Incremental cost-effectiveness			
Healthcare sector perspective			
ICER	Dominated*	Dominated*	Dominant**

	SSB Tax only	F&V Subsidy only	SSB tax to subsidize F&V
INMB	3,146,493 (39,537 to 7,810,917)	1,060,031 (-452,790 to 3,560,841)	4,217,818 (233,101 to 9,301,852)
Societal perspective			
ICER	Dominant**	Dominated***	Dominant**
INMB	8,264,522 (5,249,535 to 12,950,599)	-2,244,195 (-3,750,077 to 351,336)	6,031,350 (2,119,966 to 11,102,636)
Latino			
Prevented Healthcare Outcomes			
CHD events	15 (-3 to 40)	9 (-10 to 33)	24 (-3 to 55)
Stroke events	7 (0 to 21)	9 (0 to 24)	16 (0 to 34)
CVD deaths	4 (-6 to 17)	3 (-7 to 17)	7 (-10 to 28)
QALYs gained	6 (-3 to 19)	5 (-3 to 17)	11 (-2 to 28)
Incremental Costs (2019 USD, thousands)			
Healthcare sector perspective	-620 (-1,380 to -60)	-530 (-1,200 to -70)	-1,130 (-2,100 to -350)
Societal perspective	-3,280 (-4,320 to -2,470)	3,550 (2,700 to 4,180)	280 (-970 to 1,320)
Incremental cost-effectiveness			
Healthcare sector perspective			
ICER	Dominated*	Dominated*	Dominant**
INMB	1,707,403 (-50,143 to 4,409,486)	1,482,362 (-156,075 to 4,071,750)	3,044,140 (363,536 to 6,603,204)
Societal perspective			
ICER	Dominant**	Dominated***	$25,000
INMB	4,370,043 (2,722,188 to 7,000,592)	-2,595,430 (-4,192,981 to 34,343)	1,628,756 (-1,019,437 to 5,113,042)

Notes: All numerical results are presented as mean estimate with (95% UIs)

*Dominated (i.e., costs more and less effective) by combining SSB taxation with F&V subsidies

**Dominant (i.e., costs less and more effective) vs. status quo

***Dominated by SSB taxation alone and combining SSB taxation with F&V subsidies

Table B.12. Other Model Parameters

Parameter	Value
Number of Cycles	Until age 100 or death
Maximum number of CVD related event	2 per cycle
Sampled Size	10,000
Number of Iteration	1,000
Discount Rate	3%

APPENDIX C. SUPPLEMENTARY MATERIAL TO ESSAY 3

Table C.1. Missing percentages for each feature

Data Source	Feature Name*	Missing Percentage (%)
MDS 3.0	Gender	0.0
	Age	0.0
	Marital_Status	0.0
	Medicare_Stay	0.0
	Race	0.0
	ADL	0.0
	Cancer	0.0
	Anemia	0.0
	Artialfib: Atrial Fibrillation or Other Dysrhythmias	0.0
	CAD: Coronary Artery Disease	0.0
	DVT_PE: Deep Venous Thrombosis (DVT), Pulmonary Embolus (PE), or Pulmonary Thrombo-Embolism (PTE).	0.0
	HF: Heart Failure	0.0
	Hypertension	0.0
	Orthohypo: Orthostatic Hypotension.	0.0
	PVD_PAD: Peripheral Vascular Disease (PVD) or Peripheral Arterial Disease (PAD).	0.0
	Cirrhosis	0.0
	GERD_Ulcer: Gastroesophageal Reflux Disease (GERD) or Ulcer	0.0
	UC_CD_IBD: Ulcerative Colitis, Crohn's Disease, or Inflammatory Bowel Disease.	0.0
	BPH: Benign Prostatic Hyperplasia	0.0
	ESRD: Renal Insufficiency, Renal Failure, or End-Stage Renal Disease	0.0
	Neuroblad: Neurogenic Bladder.	0.0
	Obstrurop: Obstructive Uropathy.	0.0
	MDRO: Multidrug-Resistant Organism	0.0
	Pneumonia	0.0

Data Source	Feature Name*	Missing Percentage (%)
MDS 3.0	Septicemia	0.0
	Tuberculosis	0.0
	UTI: Urinary Tract Infection	0.0
	Viralhepatitis: Viral Hepatitis	0.0
	Woundinf: Wound Infection (other than foot).	0.0
	DM: Diabetes Mellitus (DM) (e.g., diabetic retinopathy, nephropathy, and neuropathy).	0.0
	Hyponatremia	0.0
	Hyperkalemia	0.0
	Hyperlipidemia	0.0
	Thyroiddis: Thyroid Disorder (e.g., hypothyroidism, hyperthyroidism, and Hashimoto's thyroiditis).	0.0
	Arthritis	0.0
	Osteoporosis	0.0
	Hipfracture: any hip fracture that has a relationship to current status, treatments, monitoring (e.g., sub-capital fractures, and fractures of the trochanter and femoral neck).	0.0
	Otherfracture	0.0
	Alzheimer	0.0
	Aphasia	0.0
	Cerebralpalsy	0.0
	CVA_TIA: Cerebrovascular Accident (CVA), Transient Ischemic Attack (TIA), or Stroke.	0.0
	Dementia: non-Alzheimer’s, e.g. Lewy body dementia, vascular or multi-infarct dementia; mixed dementia; frontotemporal dementia such as Pick's disease; and dementia related to stroke, Parkinson's or Creutzfeldt-Jakob diseases)	0.0
	Hemiplegia: Hemiplegia or Hemiparesis.	0.0
	Paraplegia	0.0
	Quadriplegia	0.0
	MS: Multiple Sclerosis	0.0

Data Source	Feature Name*	Missing Percentage (%)
MDS 3.0	Huntington	0.0
	Parkinson	0.0
	Tourette	0.0
	Seizure: Seizure Disorder or Epilepsy	0.0
	TBI: Traumatic Brain Injury	0.0
	Malnutrition	0.0
	Anxiety	0.0
	Depression	0.0
	Manicdepression	0.0
	Psychotic	0.0
	Schizophrenia	0.0
	PTSD: Post Traumatic Stress Disorder	0.0
	COPT_Asthma: Asthma, Chronic Obstructive Pulmonary Disease (COPD), or Chronic Lung Disease (e.g., chronic bronchitis and restrictive lung diseases such as asbestosis).	0.0
	Respiratoryfailure	0.0
	Cataracts_Glaucoma: Cataracts, Glaucoma, or Macular Degeneration	0.0
	Fracture_Hist6m: fracture related to a fall in the 6 months prior to admission/entry or reentry	0.0
	#Fall: # of falls since admission/entry	0.0
	Sbreath_Exertion: Shortness of breath or trouble breathing with exertion (e.g., walking, bathing, transferring)	0.0
	Sbreath_Rest: Shortness of breath or trouble breathing when sitting at rest.	0.0
	Sbreath_Lying: Shortness of breath or trouble breathing when lying flat	0.0
	Tabacco	0.0
	BIMS_Score	0.0
	Mood_Score	0.0
	Urinary: Urinary continence	4.2
	Bowel: Bowel continence	4.3

Data Source	Feature Name*	Missing Percentage (%)
Nursing Home Compare	Ownership	0.0
	City	0.0
	ZIP	0.0
	Bedcert: number of certified beds	0.0
	AIDHRD: Reported CNA Staffing Hours per Resident per Day	2.5
	VOCHRD: Reported LPN Staffing Hours per Resident per Day	2.5
	RNHRD: Reported RN Staffing Hours per Resident per Day	2.5
	EXP_AIDE: Expected CNA Staffing Hours per Resident per Day	0.7
	EXP_LPN: Expected LPN Staffing Hours per Resident per Day	0.7
	EXP_RN: Expected RN Staffing Hours per Resident per Day	0.7
	Cycle_1_total_score: Cycle 1 Total Health Score	0.6
	Plong_hrskulcer: Percentage of high risk long-stay residents with pressure ulcers	3.2
	Plong_pneuvaccine: Percentage of long-stay residents assessed and appropriately given the pneumococcal vaccine	2.8
	Plong_influenza: Percentage of long-stay residents assessed and appropriately given the seasonal influenza vaccine	3.0
	Plong_fall: Percentage of long-stay residents experiencing one or more falls with major injury	2.8
	Plong_depress: Percentage of long-stay residents who have depressive symptoms	2.9
	Plong_lweight: Percentage of long-stay residents who lose too much weight	2.8
	Plong_antipsychotic: Percentage of long-stay residents who received an antipsychotic medication	2.9
	Pong_pain: Percentage of long-stay residents who self-report moderate to severe pain	3.8

Data Source	Feature Name*	Missing Percentage (%)
Nursing Home Compare	Plong_phyrestrain: Percentage of long-stay residents who were physically restrained	2.8
	Plong_incADL: Percentage of long-stay residents whose need for help with daily activities has increased	3.1
	Plong_uriifect: Percentage of long-stay residents with a urinary tract infection	2.8
	Plong_catheter: Percentage of long-stay residents with a catheter inserted and left in their bladder	2.9
	Plong_lrskctrlbb: Percentage of low risk long-stay residents who lose control of their bowels or bladder	4.4
	Pshort_pneuvaccine: Percentage of short-stay residents assessed and appropriately given the pneumococcal vaccine	0.2
	Pshort_antipsychotic: Percentage of short-stay residents who newly received an antipsychotic medication	0.6
	Pshort_pain: Percentage of short-stay residents who self-report moderate to severe pain	0.7
	Pshort_influenza: Percentage of short-stay residents who were assessed and appropriately given the seasonal influenza vaccine	0.2
	Pshort_ulcer: Percentage of short-stay residents with pressure ulcers that are new or worsened	0.2
National Neighborhood Data Archive	Aland10: ZCTA land area, square miles	0.0
	Totpop13_17: Total population, ACS 2013-2017	0.0
	Popden13_17: Persons per square mile, ACS 2013-2017	0.0
	Phispanic13_17: Proportion of people of Hispanic origin, ACS 2013-2017	0.0
	Pnhwhite13_17: Proportion of people non-Hispanic White, ACS 2013-2017	0.0

Data Source	Feature Name*	Missing Percentage (%)
National Neighborhood Data Archive	Pnhblack13_17: Proportion of people non-Hispanic Black, ACS 2013-2017	0.0
	Pfborn13_17: Proportion of people who are foreign born, ACS 2013-2017	0.0
	Ped1_13_17: Proportion with Less than High School Diploma, ACS 2013-2017	0.0
	Ped2_13_17: Proportion with High School Diploma and/or Some College, ACS 2012-2013	0.0
	Ped3_13_17: Proportion with Bachelor's Degree or Higher, ACS 2013-2017	0.0
	Pin1b_13_17: Proportion of families with Income less than 15K, ACS 2013-2017	0.0
	Pin2b_13_17: Proportion of families with Income 15-30K, ACS 2013-2017	0.0
	Pin3b_13_17: Proportion of families with Income 30-50K, ACS 2013-2017	0.0
	Pin4b_13_17: Proportion of families with Income 50-100K, ACS 2013-2017	0.0
	Pin5b_13_17: Proportion of families with Income greater than 100K, ACS 2013-2017	0.0
	Pincgt75k13_17: Proportion of families with Income greater than 75K, ACS 2013-2018	0.0
	Pnvmar13_17: Proportion of People 15+ Never Married, ACS 2013-2017	0.0
	P18yr_13_17: Proportion of population under 18 yrs, ACS 2013-2017	0.0
	P18_2913_17: Proportion of population 18-29 yrs, ACS 2013-2017	0.0

Data Source	Feature Name*	Missing Percentage (%)
	P30_3913_17: Proportion of population 30-39 yrs, ACS 2013-2017	0.0
	P40_4913_17: Proportion of population 40-49 yrs, ACS 2013-2017	0.0
	P50_6913_17: Proportion of population 50-69 yrs, ACS 2013-2017	0.0
	Pge7013_17: Proportion of population 70+ yrs, ACS 2013-2017	0.0
	Punemp13_17: Proportion 16+ civ labor force unemployed, ACS 2013-2017	0.0
	Pprof13_17: Proportion emplyd civ 16+ mgmt/bus/sci/arts, ACS 2013-2017	0.0
	Ppov13_17: Proportion people w/ income past 12 months below poverty level, ACS 2013-2017	0.0
	Ppubas13_17: Proportion of households with public assistance income, ACS 2013-2017	0.0
	Pfhfam13_17: Proportion female-headed families w/ kids, ACS 2013-2017	0.0
	Pownoc13_17: Proportion owner occupied hus, ACS 2013-2017	0.0
	Disadvantage13_17: Mean of pnhblack pfhfam ppubas ppov punemp, ACS 2013-2017	0.0
	Disadvantage2_13_17: Mean of pfhfam ppubas ppov punemp, ACS 2013-2017	0.0
	Affluence13_17: Mean of pincgt75k ped3 pprof, ACS 2013-2017	0.0
	Ethnicimmigrant13_17: Mean of phispanic pfborn, ACS 2013-2017	0.0

*The features listed are before dummifying categorical variables. Also the missing percentages are calculated based on the filtered population from admission records and facilities recorded on the Nursing Home Compare dataset.

REFERENCES

1. Vogenberg FR. Predictive and Prognostic Models: Implications for Healthcare Decision-Making in a Modern Recession. *Am Health Drug Benefits*. 2009;2(6):218-222.

2. Steyerberg EW. *Clinical Prediction Models*. 1st ed. Springer-Verlag New York; 2009.

3. Wang C, Li F, Wang L, et al. The impact of population aging on medical expenses: A big data study based on the life table. *BioScience Trends*. 2017;11(6):619-631. doi:10.5582/bst.2017.01243

4. Moons KG, Royston P, Vergouwe Y. Prognosis and prognostic research: what, why, and how? *BMJ*. 2009;338:1317-1320.

5. Christensen E. Prognostic models including the Child-Pugh, MELD and Mayo risk scores—where are we and where should we go? *J Hepatol*. 2004;41:344-350.

6. Feinstein AR. Clinical judgment' revisited: the distraction of quantitative models. *Ann Intern Med*. 1994;120:799-805.

7. Concato J, Feinstein AR, Holford TR. The risk of determining risk with multivariable models. *Ann Intern Med*. 1993;118:201-210.

8. Braitman LE, Davidoff F. Predicting clinical states in individual patients. *Ann Intern Med*. 1996;125:406-412.

9. Mishra N, Silakari DS. Predictive Analytics: A Survey, Trends, Applications, Oppurtunities & Challenges. 2012;3.

10. Cohen IG, Amarasingham R, Shah A, Xie B, Lo B. The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care. *Health Affairs*. 2014;33(7):1139-1147. doi:10.1377/hlthaff.2014.0048

11. Bray J, Maxwell S. *Multivariate Analysis of Variance*. SAGE Publications, Inc.; 1985. doi:10.4135/9781412985222

12. McIntosh AR, Mišić B. Multivariate Statistical Analyses for Neuroimaging Data. *Annual Review of Psychology*. 2013;64(1):499-525. doi:10.1146/annurev-psych-113011-143804

13. Li R, Chen K, Fleisher AS, Reiman EM, Yao L, Wu X. Large-scale directional connections among multi resting-state neural networks in human brain: A functional MRI and Bayesian network modeling study. *NeuroImage*. 2011;56(3):1035-1042. doi:10.1016/j.neuroimage.2011.03.010

14. Naqishbandi TA, Ayyanathan N. Clinical Big Data Predictive Analytics Transforming Healthcare: - An Integrated Framework for Promise Towards Value Based Healthcare. In: Satapathy SC, Raju KS, Shyamala K, Krishna DR, Favorskaya MN, eds. *Advances in Decision Sciences, Image Processing, Security and Computer Vision*. Learning and Analytics in Intelligent Systems. Springer International Publishing; 2020:545-561. doi:10.1007/978-3-030-24318-0_64

15. Palanisamy V, Thirunavukarasu R. Implications of big data analytics in developing healthcare frameworks – A review. *Journal of King Saud University - Computer and Information Sciences*. 2019;31(4):415-425. doi:10.1016/j.jksuci.2017.12.007

16. Dinov ID. Methodological challenges and analytic opportunities for modeling and interpreting Big Healthcare Data. *GigaScience*. 2016;5(1):s13742-016-0117-6. doi:10.1186/s13742-016-0117-6

17. Miller T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*. 2019;267:1-38. doi:10.1016/j.artint.2018.07.007

18. Chalasani N, Younossi Z, Lavine JE, et al. The diagnosis and management of nonalcoholic fatty liver disease: Practice guidance from the American Association for the Study of Liver Diseases. *Hepatology*. 2018;67(1):328-357. doi:10.1002/hep.29367

19. Angulo P, Kleiner DE, Dam-Larsen S, et al. Liver Fibrosis, but No Other Histologic Features, Is Associated With Long-term Outcomes of Patients With Nonalcoholic Fatty Liver Disease. *Gastroenterology*. 2015;149(2):389-397.e10. doi:10.1053/j.gastro.2015.04.043

20. Vuppalanchi R, Unalp A, Natta ML. Effects of liver biopsy sample length and number of readings on sampling variability in nonalcoholic Fatty liver disease. *Clin Gastroenterol Hepatol*. 2009;7:481-486.

21. Sterling RK, Lissen E, Clumeck N. Development of a simple noninvasive index to predict significant fibrosis in patients with HIV/HCV coinfection. *Hepatology*. 2006;43:1317-1325.

22. Vuppalanchi R, Siddiqui MS, Natta ML. Performance characteristics of vibration-controlled transient elastography for evaluation of nonalcoholic fatty liver disease. *Hepatology*. 2018;67:134-144.

23. Loomba R, Cui J, Wolfson T. Novel 3D magnetic resonance elastography for the noninvasive diagnosis of advanced fibrosis in NAFLD: A prospective study. *American Journal of Gastroenterology*. 2016;111:986-994.

24. Loomba R, Wolfson T, Ang B. Magnetic resonance elastography predicts advanced fibrosis in patients with nonalcoholic fatty liver disease: a prospective study. *Hepatology*. 2014;60:1920-1928.

25. Sun WJ, Cui HL, Li N. Comparison of FIB-4 index, NAFLD fibrosis score and BARD score for prediction of advanced fibrosis in adult patients with non-alcoholic fatty liver disease: A meta-analysis study. *Hepatology Research*. 2016;46:862-870.

26. Imajo K, Kessoku T, Honda Y. Magnetic resonance imaging more accurately classifies steatosis and fibrosis in patients with nonalcoholic fatty liver disease than transient elastography. *Gastroenterology*. 2016;150:626-637.

27. Shah AG, Lydecker A, Murray K. Comparison of noninvasive markers of fibrosis in patients with nonalcoholic fatty liver disease. *Clinical Gastroenterology and Hepatology*. 2009;7:1104-1112.

28. Petta S, Vanni E, Bugianesi E. The combination of liver stiffness measurement and NAFLD fibrosis score improves the noninvasive diagnostic accuracy for severe liver fibrosis in patients with nonalcoholic fatty liver disease. *Liver Int*. 2015;35:1566-1573.

29. Tapper EB, Sengupta N, Hunink MG. Cost-effective evaluation of nonalcoholic fatty liver disease with NAFLD fibrosis score and vibration controlled transient elastography. *Am J Gastroenterol*. 2015;110:1298-1304.

30. Williams CD, Stengel J, Asike MI. Prevalence of nonalcoholic fatty liver disease and nonalcoholic steatohepatitis among a largely middle-aged population utilizing ultrasound and liver biopsy: a prospective study. *Gastroenterology*. 2011;140:124-131.

31. Scaglione S, Kliethermes S, Cao G. The epidemiology of cirrhosis in the united states: A population-based study. *J Clin Gastroenterol*. 2015;49:690-696.

32. Castera L, Foucher J, Bernard PH. Pitfalls of liver stiffness measurement: a 5-year prospective study of 13,369 examinations. *Hepatology*. 2010;51:828-835.

33. Chen J, Yin M, Talwalkar JA. Diagnostic performance of MR elastography and vibration-controlled transient elastography in the detection of hepatic fibrosis in patients with severe to morbid obesity. *Radiology*. 2017;283:418-428.

34. Yin M, Glaser KJ, Talwalkar JA. Hepatic MR elastography: Clinical performance in a series of 1377 consecutive examinations. *Radiology*. 2016;278:114-124.

35. Heimbach JK, Kulik LM, Finn RS. AASLD guidelines for the treatment of hepatocellular carcinoma. *Hepatology*. 2018;67:358-380.

36. Garcia-Tsao G, Abraldes JG, Berzigotti A. Portal hypertensive bleeding in cirrhosis: Risk stratification, diagnosis, and management: 2016 practice guidance by the American Association for the study of liver diseases. *Hepatology*. 2017;65:310-335.

37. Nakamura S, Konishi H, Kishino M. Prevalence of esophagogastric varices in patients with non-alcoholic steatohepatitis. *Hepatology Research*. 2008;38:572-579.

38. Vilar-Gomez E, Calzadilla-Bertot L, Wai-Sun Wong V, et al. Fibrosis Severity as a Determinant of Cause-Specific Mortality in Patients With Advanced Nonalcoholic Fatty Liver Disease: A Multi-National Cohort Study. *Gastroenterology*. 2018;155(2):443-457.e17. doi:10.1053/j.gastro.2018.04.034

39. Stevenson M, Lloyd-Jones M, Morgan MY. Non-invasive diagnostic assessment tools for the detection of liver fibrosis in patients with suspected alcohol-related liver disease: a systematic review and economic evaluation. *Health Technol Assess*. 2012;16:1-174.

40. Petta S, Sebastiani G, Bugianesi E. Non-invasive prediction of esophageal varices by stiffness and platelet in non-alcoholic fatty liver disease cirrhosis. *J Hepatol*. 2018;69:878-885.

41. Merli M, Nicolini G, Angeloni S. Incidence and natural history of small esophageal varices in cirrhotic patients. *J Hepatol*. 2003;38:266-272.

42. Gluud LL, Krag A. Banding ligation versus beta-blockers for primary prevention in oesophageal varices in adults. *Cochrane Database Syst Rev*. Published online 2012.

43. Cheng JW, Zhu L, Gu MJ. Meta analysis of propranolol effects on gastrointestinal hemorrhage in cirrhotic patients. *World J Gastroenterol*. 2003;9:1836-1839.

44. Imperiale TF, Chalasani N. A meta-analysis of endoscopic variceal ligation for primary prophylaxis of esophageal variceal bleeding. *Hepatology*. 2001;33:802-807.

45. Bedossa P, Dargere D, Paradis V. Sampling variability of liver fibrosis in chronic hepatitis C. *Hepatology*. 2003;38:1449-1457.

46. Singh S, Muir AJ, Dieterich DT. American gastroenterological association institute technical review on the role of elastography in chronic liver diseases. *Gastroenterology*. 2017;152:1544-1577.

47. Hsu C, Caussy C, Imajo K. Magnetic resonance vs transient elastography analysis of patients with nonalcoholic fatty liver disease: A systematic review and pooled analysis of individual participants. *Clin Gastroenterol Hepatol*. Published online 2018.

48. Majumdar A, Campos S, Gurusamy K. Defining the minimum acceptable diagnostic accuracy of noninvasive fibrosis testing in cirrhosis: A decision analytic modeling study. *Hepatology*. Published online 2019.

49. Martinez-Gonzalez MA, Ros E, Estruch R. Primary prevention of cardiovascular disease with a Mediterranean diet supplemented with extra-virgin olive oil or nuts. *N Engl J Med*. 2018;379(14):1056 1809971.

50. Mozaffarian D. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity. *Circulation*. 2016;133(2):187-225. doi:10.1161/CIRCULATIONAHA.115.018585

51. Keller A, Heitmann BL, Olsen N. Sugar-sweetened beverages, vascular risk factors and events: a systematic literature review. *Public Health Nutr*. 2015;18(07):1145-1154.

52. Xi B, Huang Y, Reilly KH. Sugar-sweetened beverages and risk of hypertension and CVD: a dose–response meta-analysis. *Br J Nutr*. 2015;113(05):709-717.

53. Wang X, Ouyang Y, Liu J, et al. Fruit and vegetable consumption and mortality from all causes, cardiovascular disease, and cancer: systematic review and dose-response meta-analysis of prospective cohort studies. *BMJ*. 2014;349:g4490. doi:10.1136/bmj.g4490

54. Dauchet L, Amouyel P, Hercberg S, Dallongeville J. Fruit and vegetable consumption and risk of coronary heart disease: a meta-analysis of cohort studies. *J Nutr*. 2006;136(10):2588-2593.

55. Rehm CD, Penalvo JL, Afshin A, Mozaffarian D. Dietary intake among US adults, 1999–2012. *JAMA*. 2016;2016;315(23):2542–53. doi:10.1001/jama.

56. Blanck HM, Gillespie C, Kimmons JE, Seymour JD, Serdula MK. Trends in fruit and vegetable consumption among US men and women, 1994–2005. *Prev Chronic Dis*. 2008;5(2). http://www.ncbi.nlm.nih.gov/pmc/

57. America HF. Compare tax policies. Healthy Food America. Published 2021. Accessed November 11, 2021. https://www.healthyfoodamerica.

58. Crosbie E, Pomeranz JL, Wright KE, Hoeper S, Schmidt L. State Preemption: An Emerging Threat to Local Sugar-Sweetened Beverage Taxation. *Am J Public Health*. 2021;111(4):677-686. doi:10.2105/AJPH.2020.306062

59. Krieger J, Bleich SN, Scarmo S, Ng SW. Sugar-Sweetened Beverage Reduction Policies: Progress and Promise. *Annu Rev Public Health*. 2021;42(1):439-461. doi:10.1146/annurev-publhealth-090419-103005

60. Moran AJ, Gu Y, Clynes S, Goheer A, Roberto CA, Palmer A. Associations between Governmental Policies to Improve the Nutritional Quality of Supermarket Purchases and Individual, Retailer, and Community Health Outcomes: An Integrative Review. *IJERPH*. 2020;17(20):7493. doi:10.3390/ijerph17207493

61. Krieger J, Magee K, Hennings T, Schoof J, Madsen KA. How sugar-sweetened beverage tax revenues are being used in the United States. *Prev Med Rep*. 2021;23(101388).

62. Engel K, Ruder EH. Fruit and Vegetable Incentive Programs for Supplemental Nutrition Assistance Program (SNAP) Participants: A Scoping Review of Program Structure. *Nutrients*. 2020;12(6):1676. doi:10.3390/nu12061676

63. Rummo PE, Lyerly R, Rose J, Malyuta Y, Cohen ED, Nunn A. The impact of financial incentives on SNAP transactions at mobile produce markets. *International Journal of Behavioral Nutrition and Physical Activity*. 2021;18(1):26. doi:10.1186/s12966-021-01093-z

64. Moran A, Thorndike A, Franckle R, et al. Financial Incentives Increase Purchases Of Fruit And Vegetables Among Lower-Income Households With Children. *Health Affairs*. 2019;38(9):1557-1566. doi:10.1377/hlthaff.2018.05420

65. Rummo PE, Noriega D, Parret A, Harding M, Hesterman O, Elbel BE. Evaluating A USDA Program That Gives SNAP Participants Financial Incentives To Buy Fresh Produce In Supermarkets. *Health Affairs*. 2019;38(11):1816-1823. doi:10.1377/hlthaff.2019.00431

66. Pearson-Stuttard J, Bandosz P, Rehm CD, et al. Comparing effectiveness of mass media campaigns with price reductions targeting fruit and vegetable intake on US cardiovascular disease mortality and race disparities. *Am J Clin Nutr*. 2017;106(1):199-206. doi:10.3945/ajcn.116.143925

67. Pearson-Stuttard J, Bandosz P, Rehm CD, et al. Reducing US cardiovascular disease burden and disparities through national and targeted dietary policies: A modelling study. *PLoS Med*. 2017;14(6):e1002311. doi:10.1371/journal.pmed.1002311

68. Jack D, Neckerman K, Schwartz-Soicher O. Socio-economic status, neighbourhood food environments and consumption of fruits and vegetables in New York City. *Public Health Nutr*. 2013;16(07):1197-1205.

69. Loftfield E, Yi SS, Curtis CJ, Bartley K, Kansagra SM. Potassium and fruit and vegetable intakes in relation to social determinants and access to produce in New York City. *Am J Clin Nutr*. 2013;98(5):1282-1288.

70. Lee-Kwan SH, Moore LV, Blanck HM, Harris DM, Galuska D. Disparities in state-specific adult fruit and vegetable consumption—United States, 2015. *MMWR Morb Mortal Wkly Rep*. 2017;66(45).

71. Jiang N, Stella SY, Russo R. Trends and sociodemographic disparities in sugary drink consumption among adults in New York City, 2009–2017. *Prev Med Rep*. 2020;19(101162).

72. Kohli-Lynch CN, Bellows BK, Thanassoulis G, et al. Cost-effectiveness of Low-density Lipoprotein Cholesterol Level-Guided Statin Treatment in Patients with Borderline Cardiovascular Risk. *JAMA Cardiology*. 2019;4(10):969-977. doi:10.1001/jamacardio.2019.2851

73. Kohli-Lynch C, Bellows B, Zhang Y, et al. Cost-Effectiveness of Lipid-Lowering Treatments in Young Adults. *Journal of the American College of Cardiology*. 2021;In press.

74. Moran AE, Odden MC, Thanataveerat A, et al. Cost-effectiveness of hypertension therapy according to 2014 guidelines. *N Engl J Med*. 2015;372(5):447-455. doi:10.1056/NEJMsa1406751

75. Weinstein MC, Coxson PG, Williams LW, Pass TM, Stason WB, Goldman L. Forecasting coronary heart disease incidence, mortality, and cost: the Coronary Heart Disease Policy Model. *Am J Public Health*. 1987;77(11):1417-1426. doi:10.2105/ajph.77.11.1417

76. Zhang Y, Vittinghoff E, Pletcher MJ, et al. Associations of Blood Pressure and Cholesterol Levels During Young Adulthood With Later Cardiovascular Events. *Journal of the American College of Cardiology*. 2019;74(3):330-341. doi:10.1016/j.jacc.2019.03.529

77. Oelsner EC, Balte PP, Cassano PA, et al. Harmonization of Respiratory Data from 9 US Population-Based Cohorts. *American Journal of Epidemiology*. 2018;187(11):2265-2278. doi:10.1093/aje/kwy139

78. Zeki Al Hazzouri A, Vittinghoff E, Zhang Y, et al. Use of a pooled cohort to impute cardiovascular disease risk factors across the adult life course. *International Journal of Epidemiology*. 2019;48(3):1004-1013. doi:10.1093/ije/dyy264

79. *Community Health Survey 2018 Public Use Dataset*. New York City Department of Health and Mental Hygiene.; 2018. Accessed April 4, 2021. https://www1.nyc.gov/site/doh/data/data-sets/community-health-survey-public-use-data.page

80. Afshin A, Peñalvo JL, Gobbo LD, et al. The prospective impact of food pricing on improving dietary consumption: A systematic review and meta-analysis. *PLOS ONE*. 2017;12(3):e0172277. doi:10.1371/journal.pone.0172277

81. Narain A, Kwok CS, Mamas MA. Soft drinks and sweetened beverages and the risk of cardiovascular disease and mortality: a systematic review and meta-analysis. *International Journal of Clinical Practice*. 2016;70(10):791-805. doi:10.1111/ijcp.12841

82. Micha R, Peñalvo JL, Cudhea F, Imamura F, Rehm CD, Mozaffarian D. Association between dietary factors and mortality from heart disease, stroke, and type 2 diabetes in the United States. *JAMA - Journal of the American Medical Association*. 2017;317(9):912-924. doi:10.1001/jama.2017.0947

83. Moran AE, Forouzanfar MH, Roth GA, et al. Temporal trends in ischemic heart disease mortality in 21 world regions, 1980 to 2010: The global burden of disease 2010 study. *Circulation*. 2014;129(14):1483-1492. doi:10.1161/CIRCULATIONAHA.113.004042

84. Moran AE, Forouzanfar MH, Roth GA. The global burden of ischemic heart disease in 1990 and 2010: the Global Burden of Disease 2010 study. 2014;129:1493-1501. doi:10.1161/CIRCU

85. Murray CJL, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990-2010: A systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2012;380(9859):2197-2223. doi:10.1016/S0140-6736(12)61689-4

86. CDC WONDER Online Database Underlying Cause of Death 2005. Centers for Disease Control and Prevention. Accessed April 4, 2021. https://wonder.cdc.gov/

87. Sanders GD, Neumann PJ, Basu A, et al. Recommendations for Conduct, Methodological Practices, and Reporting of Cost-effectiveness Analyses: Second Panel on Cost-Effectiveness in Health and Medicine. *JAMA*. 2016;316(10):1093-1103. doi:10.1001/jama.2016.12195

88. Eze-Nliam CM, Zhang Z, Weiss SA, Weintraub WS. Cost-effectiveness Assessment of Cardiac Interventions: Determining a Socially Acceptable Cost Threshold. *Interv Cardiol (Lond)*. 2014;6(1):45-55. doi:10.2217/ica.13.81

89. Lee Y, Mozaffarian D, Sy S, et al. Cost-effectiveness of financial incentives for improving diet and health through Medicare and Medicaid: A microsimulation study. *PLOS Medicine*. 2019;16(3):e1002761. doi:10.1371/journal.pmed.1002761

90. Lee Y, Mozaffarian D, Sy S, et al. Health impact and cost-effectiveness of volume, tiered, and absolute sugar content sugar-sweetened beverage tax policies in the United States: a micro-simulation study. *Circulation*. 2020;142(6):523-534. doi:10.1161/CIRCULATIONAHA.119.042956

91. Purtle J, Langellier B, Lê-Scherban F. A Case Study of the Philadelphia Sugar-Sweetened Beverage Tax Policymaking Process: Implications for Policy Development and Advocacy. *J Public Health Manag Pract*. 2018;24(1):4-8. doi:10.1097/PHH.0000000000000563

92. Jou J, Niederdeppe J, Barry CL, Gollust SE. Strategic messaging to promote taxation of sugar-sweetened beverages: lessons from recent political campaigns. *Am J Public Health*. 2014;104(5):847-853. doi:10.2105/AJPH.2013.301679

93. Long MW, Polacsek M, Bruno P. Cost-Effectiveness Analysis and Stakeholder Evaluation of 2 Obesity Prevention Policies in Maine, US. *J Nutr Educ Behav*. 2019;51(10):1177-1187. doi:10.1016/j.jneb.2019.07.005

94. Teng AM, Jones AC, Mizdrak A, Signal L, Genç M, Wilson N. Impact of sugar-sweetened beverage taxes on purchases and dietary intake: systematic review and meta-analysis. *Obes Rev*. 2019;20(9):1187-1204.

95. Popkin BM, Ng SW. Sugar-sweetened beverage taxes: lessons to date and the future of taxation. *PLoS Med*. 2021;18(1).

96. Shi Z, Ruel G, Dal Grande E, Pilkington R, Taylor AW. Soft drink consumption and multimorbidity among adults. *Clin Nutr ESPEN*. 2015;10(2).

97. Long MW, Gortmaker SL, Ward ZJ. Cost effectiveness of a sugar-sweetened beverage excise tax in the US. *Am J Prev Med*. 2015;49(1):112-123.

98. Good Health, Good Value: NYC Receives $5.5 Million Grant to Make Healthy Food More Affordable - NYC Health. Accessed August 23, 2021. https://www1.nyc.gov/site/doh/about/press/pr2021/good-health-good-value-nyc-receives-grant-for-affordable-healthy-food.page

99. Payne GH, Wethington H, Olsho L, Jernigan J, Farris R, Walker DK. Peer reviewed: implementing a farmers' market incentive program: perspectives on the new york city health bucks program. Prev chronic dis. 2013;10.

100. The Obesity Evidence Hub. Countries that have taxes on sugar-sweetened beverages (SSBs. Cancer Council Victoria. Published 2022. www.obesityevidencehub.org.au.

101. Crosbie E, Florence D. Expanding our understanding of industry opposition to help implement sugar-sweetened beverage taxation. *Public Health Nutr Published online*. 2021:1-3.

102. Pomeranz JL, Pertschuk M. State preemption: a significant and quiet threat to public health in the United States. *Am J Public Health*. 2017;107(6):900-902.

103. Crosbie E, Schillinger D, Schmidt LA. State preemption to prevent local taxation of sugar-sweetened beverages. *JAMA Intern Med*. 2019;179(3):291-292.

104. Falbe J, Madsen K. Growing Momentum for Sugar-Sweetened Beverage Campaigns and Policies: Costs and Considerations. *Am J Public Health*. 2017;107(6):835-838. doi:10.2105/AJPH.2017.303805

105. Romanos-Nanclares A, Toledo E, Gardeazabal I, Jimé-nez-Moleón JJ, Martínez-González MA, Gea A. Sugar-sweetened beverage consumption and incidence of breast cancer: the Seguimiento Universidad de Navarra (SUN) Project. *Eur J Nutr*. 2019;58(7):2875-2886.

106. Terry P, Terry JB, Wolk A. Fruit and vegetable consumption in the prevention of cancer: an update. *J Intern Med*. 2001;250(4):280-290.

107. Carter C. The Need to Reform Medicare's Payments to Skilled Nursing Facilities is as Strong as Ever. Published online 2015. Accessed January 23, 2023. https://www.urban.org/sites/default/files/publication/39036/2000072-The-Need-to-Reform-Medicare-Payments-to-SNF.pdf

108. Commission MPA. Data book: Health care spending and the medicare program. Published online 2017.

109. Chandra A, Dalton MA, Holmes J. Large increases in spending on postacute care in Medicare point to the potential for cost savings in these settings. *Health Affairs*. 2013;32(5):864-872.

110. Xu D, Kane R, Arling G. Relationship between nursing home quality indicators and potentially preventable hospitalisation. *BMJ Qual Saf*. 2019;28(7):524-533. doi:10.1136/bmjqs-2018-008924

111. Rosen BT, Halbert RJ, Hart K, Diniz MA, Isonaka S, Black JT. The Enhanced Care Program: Impact of a Care Transition Program on 30-Day Hospital Readmissions for Patients Discharged From an Acute Care Facility to Skilled Nursing Facilities. *Journal of Hospital Medicine*. 2018;13(4):229-235. doi:10.12788/jhm.2852

112. Mor V, Intrator O, Feng Z, Grabowski DC. The revolving door of rehospitalization from skilled nursing facilities. *Health Affairs*. 2010;29(1):57-64. doi:10.1377/hlthaff.2009.0629

113. Hines A, Barrett M, Jiang HJ, Steiner C. Conditions with the largest number of adult hospital readmissions by payer, 2011: statistical brief# 172. Published online April 2014.

114. Burke RE, Whitfield EA, Hittle D, et al. Hospital Readmission From Post-Acute Care Facilities: Risk Factors, Timing, and Outcomes. *Journal of the American Medical Directors Association*. 2016;17(3):249-255. doi:10.1016/j.jamda.2015.11.005

115. Li Y, Cai X, Yin J, Glance LG, Mukamel DB. Is Higher Volume of Postacute Care Patients Associated With a Lower Rehospitalization Rate in Skilled Nursing Facilities? *Med Care Res Rev*. 2012;69(1):103-118. doi:10.1177/1077558711414274

116. Tomlinson J, Cheong VL, Fylan B, et al. Successful care transitions for older people: a systematic review and meta-analysis of the effects of interventions that support medication continuity. *Age Ageing*. 2020;49(4):558-569. doi:10.1093/ageing/afaa002

117. Grabowski DC, Joynt Maddox KE. Postacute Care Preparedness for COVID-19: Thinking Ahead. *JAMA*. 2020;323(20):2007-2008. doi:10.1001/jama.2020.4686

118. Gettel CJ, Voils CI, Bristol AA, et al. Care transitions and social needs: A Geriatric Emergency care Applied Research (GEAR) Network scoping review and consensus statement. *Acad Emerg Med*. 2021;28(12):1430-1439. doi:10.1111/acem.14360

119. Hospital Readmissions Reduction Program (HRRP) | CMS. Accessed February 19, 2021. https://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/Readmissions-Reduction-Program

120. Kangovi S, Grande D. Hospital readmissions—not just a measure of quality. *Jama*. 2011;306(16):1796-1797.

121. Unroe KT, Carnahan JL, Hickman SE, Sachs GA, Hass Z, Arling G. The Complexity of Determining Whether a Nursing Home Transfer Is Avoidable at Time of Transfer. *J Am Geriatr Soc*. 2018;66(5):895-901. doi:10.1111/jgs.15286

122. Unroe KT, Hickman SE, Carnahan JL, Hass Z, Sachs G, Arling G. Investigating the Avoidability of Hospitalizations of Long Stay Nursing Home Residents: Opportunities for Improvement. *Innov Aging*. 2018;2(2):igy017. doi:10.1093/geroni/igy017

123. Bogaisky M, Dezieck L. Early Hospital Readmission of Nursing Home Residents and Community-Dwelling Elderly Adults Discharged from the Geriatrics Service of an Urban Teaching Hospital: Patterns and Risk Factors. *Journal of the American Geriatrics Society*. 2015;63(3):548-552. doi:10.1111/jgs.13317

124. Murray F, Allen M, Clark CM, Daly CJ, Jacobs DM. Socio-demographic and -economic factors associated with 30-day readmission for conditions targeted by the hospital readmissions reduction program: a population-based study. *BMC Public Health*. 2021;21(1):1922. doi:10.1186/s12889-021-11987-z

125. Gupta S, Zengul FD, Davlyatov GK, Weech-Maldonado R. Reduction in Hospitals' Readmission Rates: Role of Hospital-Based Skilled Nursing Facilities. *Inquiry*. 2019;56:46958018817994. doi:10.1177/0046958018817994

126. Neuman MD, Wirtalla C, Werner RM. Association Between Skilled Nursing Facility Quality Indicators and Hospital Readmissions. *JAMA*. 2014;312(15):1542-1551. doi:10.1001/jama.2014.13513

127. Chandra A, Rahman PA, Sneve A, et al. Risk of 30-Day Hospital Readmission Among Patients Discharged to Skilled Nursing Facilities: Development and Validation of a Risk-Prediction Model. *Journal of the American Medical Directors Association*. 2019;20(4):444-450.e2. doi:10.1016/j.jamda.2019.01.137

128. Kimball CC, Nichols CI, Nunley RM, Vose JG, Stambough JB. Skilled Nursing Facility Star Rating, Patient Outcomes, and Readmission Risk After Total Joint Arthroplasty. *The Journal of Arthroplasty*. 2018;33(10):3130-3137. doi:10.1016/j.arth.2018.06.020

129. Silverstein MD, Qin H, Mercer SQ, Fong J, Haydar Z. Risk factors for 30-day hospital readmission in patients ≥65 years of age. *Proc (Bayl Univ Med Cent)*. 2008;21(4):363-372. doi:10.1080/08998280.2008.11928429

130. Howard EP, Morris JN, Schachter E, Schwarzkopf R, Shepard N, Buchanan ER. Machine-Learning Modeling to Predict Hospital Readmission Following Discharge to Post-Acute Care. *Journal of the American Medical Directors Association*. 2021;22(5):1067-1072.e29. doi:10.1016/j.jamda.2020.12.017

131. Nursing homes including rehab services datasets. Accessed March 10, 2022. https://data.cms.gov/provider-data/search?theme=Nursing%20homes%20including%20rehab%20services

132. Melendez R, Clarke P, Khan A, Gomez-Lopez I, Li M, Chenoweth M. National Neighborhood Data Archive (NaNDA): Socioeconomic Status and Demographic Characteristics of ZIP Code Tabulation Areas, United States, 2008-2017. Published online 2020. doi:10.3886/E120462

133. Buuren S van, Groothuis-Oudshoorn K. mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*. 2011;45:1-67. doi:10.18637/jss.v045.i03

134. Batista GEAPA, Prati RC, Monard MC. A study of the behavior of several methods for balancing machine learning training data. *SIGKDD Explor Newsl*. 2004;6(1):20-29. doi:10.1145/1007730.1007735

135. Buda M, Maki A, Mazurowski MA. A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*. 2018;106:249-259. doi:10.1016/j.neunet.2018.07.011

136. Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*. 2002;16:321-357. doi:10.1613/jair.953

137. Wilson DL. Asymptotic Properties of Nearest Neighbor Rules Using Edited Data. *IEEE Transactions on Systems, Man, and Cybernetics*. 1972;SMC-2(3):408-421. doi:10.1109/TSMC.1972.4309137

138. Tibshirani R. Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1996;58(1):267-288. doi:10.1111/j.2517-6161.1996.tb02080.x

139. Polson NG, Scott JG, Windle J. Bayesian Inference for Logistic Models Using Pólya–Gamma Latent Variables. *Journal of the American Statistical Association*. 2013;108(504):1339-1349. doi:10.1080/01621459.2013.829001

140. Nelder JA, Wedderburn RWM. Generalized Linear Models. *Journal of the Royal Statistical Society: Series A (General)*. 1972;135(3):370-384. doi:10.2307/2344614

141. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32. doi:10.1023/A:1010933404324

142. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. Vol 13-17-Augu. ; 2016:785-794. doi:10.1145/2939672.2939785

143. Bavarian B. Introduction to neural networks for intelligent control. *IEEE Control Systems Magazine*. 1988;8(2):3-7. doi:10.1109/37.1866

144. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. NIPS'17. Curran Associates Inc.; 2017:4768-4777.

145. Shapley LS. 17. A value for n-Person games. In: Kuhn HW, Tucker AW, eds. *Contributions to the Theory of Games (AM-28), Volume II*. Princeton University Press; 1953:307-318. doi:doi:10.1515/9781400881970-018

146. Leider J, Powell LM. Sugar-sweetened beverage prices: Variations by beverage, food store, and neighborhood characteristics, 2017. *Preventive Medicine Reports*. 2019;15:100883. doi:10.1016/j.pmedr.2019.100883

147. Andreyeva T, Long MW, Brownell KD. The impact of food prices on consumption: A systematic review of research on the price elasticity of demand for food. *American Journal of Public Health*. 2010;100(2):216-222. doi:10.2105/AJPH.2008.151415

148. *Supplementary Nutrition Assistance Program Participation and Costs, 1969-2019*. Economic Research Service (ERS). U.S. Department of Agriculture (USDA) Accessed August 10, 2021. https://fns-prod.azureedge.net/sites/default/files/resource-files/SNAPsummary-7.pdf

149. Bartlett S, Klerman J, Olsho L, et al. *Evaluation of the Healthy Incentives Pilot (HIP): Final Report*. Prepared by Abt Associates for the U.S. Department of Agriculture, Food and Nutrition Service; 2014.

150. Goldberg RJ, McCormick D, Gurwitz JH, Yarzebski J, Lessard D, Gore JM. Age-related trends in short- and long-term survival after acute myocardial infarction: a 20-year population-based perspective (1975-1995). *Am J Cardiol*. 1998;82(11):1311-1317.

151. Krumholz HM, Chen J, Chen YT, Wang Y, Radford MJ. Predicting one-year mortality among elderly survivors of hospitalization for an acute myocardial infarction: results from the Cooperative Cardiovascular Project. *J Am Coll Cardiol*. 2001;38(2):453-459.

152. Rogers WJ, Canto JG, Barron HV, Boscarino JA, Shoultz DA, Every NR. Treatment and outcome of myocardial infarction in hospitals with and without invasive capability. Investigators in the National Registry of Myocardial Infarction. *J Am Coll Cardiol*. 2000;35(2):371-379.

153. Vaccarino V, Krumholz HM, Yarzebski J, Gore JM, Goldberg RJ. Sex differences in 2-year mortality after hospital discharge for myocardial infarction. *Ann Intern Med*. 2001;134(3):173-181.

154. Canto JG, Rogers WJ, Chandra NC, et al. The Association of Sex and Payer Status on Management and Subsequent Survival in Acute Myocardial Infarction. *Arch Intern Med*. 2002;162(5):587-593. doi:10.1001/archinte.162.5.587

155. Lampe FC, Whincup PH, Wannamethee SG, Shaper AG, Walker M, Ebrahim S. The natural history of prevalent ischaemic heart disease in middle-aged men. *Eur Heart J*. 2000;21(13):1052-1062. doi:10.1053/euhj.1999.1866

156. Witt BJ, Brown RD, Jacobsen SJ, Weston SA, Yawn BP, Roger VL. A community-based study of stroke incidence after myocardial infarction. *Ann Intern Med*. 2005;143(11):785-792.

157. Lakshminarayan K, Schissel C, Anderson DC, et al. Five-year rehospitalization outcomes in a cohort of patients with acute ischemic stroke: Medicare linkage study. *Stroke*. 2011;42(6):1556-1562. doi:10.1161/STROKEAHA.110.605600

158. Anderson CS, Jamrozik KD, Broadhurst RJ, Stewart-Wynne EG. Predicting survival for 1 year among different subtypes of stroke. Results from the Perth Community Stroke Study. *Stroke*. 1994;25(10):1935-1944.

159. Appelros P, Gunnarsson KE, Terént A. Ten-year risk for myocardial infarction in patients with first-ever stroke: a community-based study. *Acta Neurol Scand*. 2011;124(6):383-389. doi:10.1111/j.1600-0404.2011.01500.x

160. National Hospital Discharge Suvey. Published online 2000. https://www.cdc.gov/nchs/nhds/index.htm

161. *Healthcare Quality and Analysis Division, Report on Heart Attack Outcomes in California 1996-1998. Volume 1: User's Guide*. California Office of Statewide Health Planning and Development; 2002.

162. *Healthcare Quality and Analysis Division, Report on Heart Attack Outcomes in California 1996-1998. Volume 2: Technical Guide*. California Office of Statewide Health Planning and Development; 2002.

163. Patient Discharge Data, Version A. 1998 and 2000. Published online 2000.

164. CDC Wonder: Stroke Mortality Rates. Published online 2005.

165. *Community Health Survey 2018 Public Use Dataset*.; 2018.

166. *The Cost of the Recommended Daily Servings of Fresh Produce Produce Marketing Association*.; 2010.

167. Leider J, Powell LM. Sugar-sweetened beverage prices: Variations by beverage, food store, and neighborhood characteristics, 2017. *Preventive Medicine Reports*. 2019;15:100883. doi:10.1016/j.pmedr.2019.100883

168. Andreyeva T, Chaloupka FJ, Brownell KD. Estimating the potential of taxes on sugar-sweetened beverages to reduce consumption and generate revenue. *Preventive Medicine*. 2011;52(6):413-416. doi:10.1016/j.ypmed.2011.03.013

169. Micha R, Peñalvo JL, Cudhea F, Imamura F, Rehm CD, Mozaffarian D. Association between dietary factors and mortality from heart disease, stroke, and type 2 diabetes in the United States. *JAMA - Journal of the American Medical Association*. 2017;317(9):912-924. doi:10.1001/jama.2017.0947

170. Mozaffarian D. Dietary and Policy Priorities for Cardiovascular Disease, Diabetes, and Obesity. *Circulation*. 2016;133(2):187-225. doi:10.1161/CIRCULATIONAHA.115.018585

171. Narain A, Kwok CS, Mamas MA. Soft drinks and sweetened beverages and the risk of cardiovascular disease and mortality: a systematic review and meta-analysis. *International Journal of Clinical Practice*. 2016;70(10):791-805. doi:10.1111/ijcp.12841

172. Medical Expenditure Panel Survey public use files 1998-2008. Published online 2008. http://meps.ahrq.gov/mepsweb/

173. *California Public Patient Discharge Data*. California Office of Statewide Health Planning and Development.; 2008. https://data.bls.gov/timeseries/CUUR0000SAM?output_view=pct_12mths

174. *Hospital Financial Data*. California Office of Statewide Health Planning and Development.; 2008. https://data.bls.gov/timeseries/CUUR0000SAM?output_view=pct_12mths

175. Bureau of Labor Statistics Data. Consumer Price Index for all urban consumers. data.bls.gov. Published January 2019. Accessed October 24, 2018. https://data.bls.gov/timeseries/CUUR0000SAM?output_view=pct_12mths

176. Moran AE, Forouzanfar MH, Roth GA, et al. Temporal trends in ischemic heart disease mortality in 21 world regions, 1980 to 2010: the Global Burden of Disease 2010 study. *Circulation*. 2014;129(14):1483-1492. doi:10.1161/CIRCULATIONAHA.113.004042

177. Moran AE, Forouzanfar MH, Roth GA, et al. The global burden of ischemic heart disease in 1990 and 2010: the Global Burden of Disease 2010 study. *Circulation*. 2014;129(14):1493-1501. doi:10.1161/CIRCULATIONAHA.113.004046

178. Murray CJL, Vos T, Lozano R, et al. Disability-adjusted life years (DALYs) for 291 diseases and injuries in 21 regions, 1990–2010: a systematic analysis for the Global Burden of Disease Study 2010. *The Lancet*. 2012;380(9859):2197-2223. doi:10.1016/S0140-6736(12)61689-4