Three Essays in Health Care

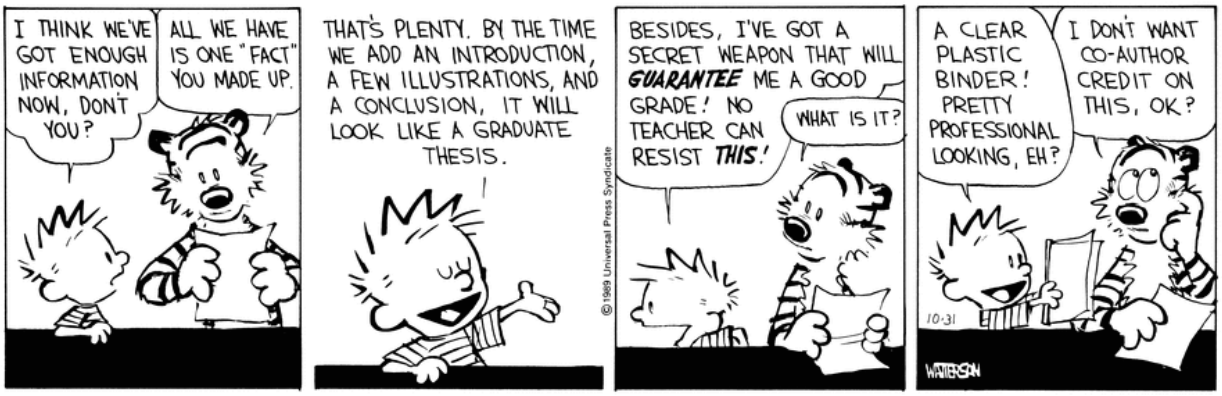
By

Anup Das

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Health Services Organization and Policy)
in the University of Michigan
2016

Doctoral Committee:

Professor Richard A. Hirth, Co-Chair
Professor Lisa A. Prosser, Co-Chair
Professor William H. Herman
Professor Edward C. Norton



Dedication

I dedicate this dissertation to my grandfather, Shakyasingh Das. A schoolteacher in a remote village in Eastern India, my grandfather dedicated his life to teaching, no matter how scarce the resources. He made sure my father received the best education possible, so that my dad could go from being a student in a small village in India to a Distinguished Professor at a university in the United States. This is my attempt at continuing the legacy my grandfather started.

Acknowledgements

First and foremost, I would like to thank my committee for their continued support of my learning. Dr. Hirth, your support in helping me contextualize my research projects and your ability to identify strengths and weaknesses in research designs has been invaluable over the last couple of years. I enjoyed our ability to mix sports and health economics during our advising sessions, and hope we can continue to do so in the future. Dr. Prosser, your willingness to regularly meet with me and your positivity provided immeasurable support to me during my PhD years. Your guidance for my research as well as professional goals has been instrumental in getting me to this position, and I truly appreciate your advising. Dr. Herman, your universal knowledge of diabetes was crucial at every stage of my research projects, and I cannot thank you enough for all the valuable feedback you have provided to me over the past couple of years. Dr. Norton, I continue to learn so much from you both inside and outside of the classroom and I don't think I could have asked for a better professor to teach me my last class in my 20th year of schooling.

I would also like to thank Dr. Wen Ye and the Michigan Center for Diabetes Translational Research for their help with my research projects at every step. Dr. Ye, thank you for walking me through the Michigan Model for Diabetes, as I would not have been able to navigate the program without your help.

Although a PhD is meant to be individual research, none of this dissertation was possible without the continued support of so many people. I would like to thank Yubraj Acharya, Betsy Cliff, Morris Hamilton, and Qing Zheng for their patience in accommodating my every question about econometrics over the past four years. Without them, I would not know what an identification strategy is. I also would like to thank Michael Harvey for all his help in teaching me decision sciences and agreeing to review all of my documents for my grant and decision science papers from day 1. In addition, many of my other friends outside of the program have provided considerable support in every stage of this process, and for that I am thankful.

I also would like to thank the NIH and the National Institute of Diabetes and Digestive and Kidney Diseases for their funding support (Award No. F30DK105749).

Finally, I would like to thank my family, Mom, Dad, Bhai, Bhauja, Saina, Bittu, Murali Uncle, and Radha Aunty for their continued support of my education. You help me take the necessary mental breaks from work to enjoy great food and laughter, and your love is always appreciated. Finally, thank you to Preeti for putting up with me being a professional graduate student for life. Your love and support make all this work enjoyable, and I promise I will graduate at some point.

Table of Contents

List of Tables

List of Figures

Abstract

This dissertation examines two common sources of increased health care costs – readmissions and the co-occurrence of depression among patients with diabetes. The first paper examines hospital performance in the Hospital Readmissions Reduction Program to determine whether sources of incentive heterogeneity are associated with differences in improvements over multiple years. I find that hospitals seem to be responding to the main incentive in the program, as those that performed poorly in previous years improve significantly more than hospitals that have avoided penalties. Hospitals also are making improvements in conditions that have the highest marginal benefit from better performance. Payer mix does not seem to be correlated with hospital performance over time even though the financial incentives of the program only apply to future Medicare reimbursements. In the second paper I develop a model to predict the onset of depression among individuals with diabetes. Using data from the Health and Retirement Study and the National Health and Nutrition Examination Survey, I find that gender, body-mass index, hypertension, history of stroke, history of heart disease, and duration of diabetes are significant predictors of annual depression status. I then build this depression prediction algorithm into the Michigan Model for Diabetes, an existing microsimulation model that allows users to evaluate the progression of diabetes. In the final paper, I use the modified diabetes simulation model to evaluate the cost-effectiveness of the collaborative care intervention to treat depression among patients with diabetes. Trials suggest that the collaborative care intervention, a multidisciplinary

approach to address the depressive symptoms of patients, can be cost-effective in the short-term when used to treat patients with diabetes and comorbid depression. Using simulation models allows us to evaluate the long-term cost-effectiveness as well as the influence of a variety of inputs on the value of the program. Only when the utility loss associated with depression is small or the intervention effectiveness is substantially decreased does the intervention require a higher willingness-to-pay to be considered cost-effective. Otherwise, our base-case analysis and other one-way sensitivity analyses support the conclusion that this intervention is cost-effective.

Chapter 1
Introduction

Every year there is more news surrounding the costs of healthcare in the United States, and the lower than expected quality of care. We pay more than many other countries, without reaping a parallel shift in many common indicators of health care quality. It is within this context that this dissertation examines two aspects of health and health care. In Chapter 2, we analyze the Hospital Readmissions Reduction Program, an incentive program stemming from the Affordable Care Act with the purpose of incentivizing hospitals to reduce their preventable readmissions across a number of common diagnoses. Readmissions can account for a sizeable share of health care spending, and research suggests that many readmissions could be prevented. Using a uniform methodology to evaluate hospital performance and determine the size of financial penalties, the Centers for Medicare & Medicaid Services has been penalizing hospitals for four years now. We were interested in determining whether various sources of incentive heterogeneity in the program were associated with improvements over time. We find that hospital program performance in previous years and marginal benefits of improvement correlate strongly with changes in hospital readmissions over time. On the other hand, the proportion of Medicare patients and thus the share of financial reimbursements at stake do not seem to be associated with changes in hospital performance over time. As hospitals continue to face these financial incentives to improve, it may become more important for them to identify the ways to most efficiently participate in the Hospital Readmissions Reduction Program.

Another source of health care spending and quality concern exists within the diabetes population. A chronic illness that is associated with numerous complications and comorbid diseases, diabetes severely impacts the health of millions of individuals and inevitably leads to increased health care spending. The risk of depression is twice as high in patients with diabetes compared to the general population. When an individual develops both depression and diabetes, they experience increased risks of developing the complications and comorbidities associated with diabetes, suffer increased health care costs, and score lower on health utility scales. Therefore, identifying viable treatment modalities to address the health care needs of this population could vastly improve their health and reduce their health care costs. One treatment approach that has gained traction in recent years is the collaborative care strategy. This intervention combines physicians, nurses and/or case managers with patients to use active follow-up and case monitoring over a 12-month time period to improve depressive symptoms. Evidence suggests that this treatment strategy has health and economic benefits for up to 2 years, but evidence from modeling studies could improve the limited knowledge surrounding this intervention.

In Chapter 3 we build a model to predict the development of depression among patients with diabetes using data from the Health and Retirement Study and the National Health and Nutrition Examination Survey. Using a random-effects logistic model, we predict individual depression status based on gender, body-mass index, hypertension, history of stroke, history of heart disease, and duration of diabetes. This prediction model is moderately discriminatory. We then build this into the Michigan Model for Diabetes, an existing model that simulates the progression of diabetes among a cohort of individuals, so that researchers can use microsimulation to study treatment options for patients with both depression and diabetes.

In Chapter 4, we use our modifications to the Michigan Model for Diabetes developed in Chapter 3 to study the cost-effectiveness of the collaborative care intervention to treat depression among patients with diabetes. Using data from existing studies and a variety of assumptions to support our model inputs, we study the short and long-term health and economic benefits of the collaborative care approach. The results from our model suggest that the collaborative care intervention can be very cost-effective and a high value investment. In many periods, the intervention dominates the usual care setting. Only when the utility loss associated with depression is small or the effectiveness of the intervention is minimized does the cost-effectiveness of this approach require higher willingness-to-pay. Under our base-case scenario and a variety of other one-way sensitivity analyses, we find that a relatively small willingness-to-pay would render the collaborative care intervention cost-effective.

Chapter 2
Hospital Responses to Incentive Heterogeneity in the Hospital Readmissions Reduction Program

Introduction

Unnecessary readmissions continue to be a major health care problem in the United States, and can lead to significantly increased health care costs (Goldfield et al., 2008). In a fragmented health care delivery system where patients may be discharged too early, given few instructions on what to do at home, or discharged without proper coordination with their outpatient providers, preventable readmissions have increasingly come under the spotlight as an area for immediate improvement (Bartel, 2014; Berwick and Hackbarth, 2012). For many decades, readmissions have been studied as a potential marker of poor inpatient care quality as well as a source for increased health care costs (Anderson and Steinberg, 1984; Ashton et al., 1995). In 2011, overall readmission rates were the highest for the Medicare population, with 17.2% of patients readmitted within 30-days (Hines, 2014). Currently, Medicare alone spends more than $17.8 billion a year on avoidable readmissions (Ness, 2013). While not all readmissions are necessarily indicators of low quality care, a sizeable share of them could be preventable (Friedman and Basu, 2004).

The Hospital Readmissions Reduction Program (HRRP) began penalizing hospitals for their excess readmissions across a variety of conditions in fiscal year 2013. After four years, the share of hospitals receiving penalties has increased, but evidence suggests that hospitals have improved their readmissions performance (Boccuti, 2015). For all participating hospitals in this

program, the Centers for Medicare & Medicaid Services (CMS) uses a uniform methodology to determine whether or not they will receive a penalty, as well as the size of any penalties. CMS determines the proportion of excess readmissions by comparing individual hospital performance to an average hospital's expected performance, and uses this to determine the amount of money spent by each hospital on these preventable readmissions. The excess payment amount, divided by a hospital's total DRG reimbursements represents the proportion of their reimbursements spent on excess readmissions, which is translated into a penalty. Despite the uniformity of the methodology, there exist many sources of heterogeneity in the real incentives individual hospitals face each fiscal year. We are interested in determining the extent to which some of the primary sources of heterogeneity in the incentives of the Hospital Readmission Reduction Program are associated with improvement in hospital performance over time.

With incentive programs where performance drives financial implications, there will be inherent variation in the incentives hospitals face. Baseline performance levels, importance placed on avoiding financial penalties, importance placed on public perception of quality, amount of financial revenue at risk, importance of specific condition service lines in hospital care portfolio, and ability to engage in quality improvement efforts all are potential factors that could alter whether and how well hospitals respond to this type of program. Some hospitals avoid penalties, while for those receiving penalties, the size of the penalty depends on calculated program performance. Thus, there is heterogeneity in the size as well as the overall incentives to improve readmissions performance, so we expect to see differences in how hospitals respond to the program. The contribution of this analysis is to identify and understand how these heterogeneous incentives may be correlated with hospital program performance over time. The incentive program applies the same financial penalty structure to hospitals irrespective of the

number of Medicare patients they have, the share of patients in their hospital who have primary diagnoses that are one of the program conditions, and whether or not hospitals are making improvements in their readmission performance. Further, with the data available, we develop a method to calculate a standardized marginal improvement in condition-specific performance for participating hospitals. Using this standardized improvement, we are then able to calculate the marginal benefit of condition-specific improvement on program performance. Finally, we determine if hospitals are improving their readmissions performance in conditions with the greatest marginal benefit on overall program performance.

It is under the assumption that a proportion of readmissions are avoidable that CMS began the Hospital Readmissions Reduction Program (HRRP). After the passing of the Affordable Care Act (ACA), CMS was able to begin three programs that would reimburse hospitals based on the value and quality of services they provided in addition to their volume. To incentivize hospitals to work on reducing their preventable readmission rates, the HRRP reduces Medicare reimbursements to acute care hospitals with excess readmissions in a defined set of conditions. Readmissions in this program are defined as an admission to a hospital within 30 days of discharge, and CMS calculates the proportion of readmissions that are excess for all hospitals. The reduction in Medicare reimbursements is levied as a percentage point decrease in a hospital's base-operating diagnosis-related group (DRG) amount for that fiscal year of the program (Centers for Medicare & Medicaid Services, 2013). This is the base reimbursement amount that is adjusted for geographic factors, before any policy, case-mix, or transfer adjustments take place in determining how much a hospital will be reimbursed. CMS has chosen several high-volume conditions to ensure that hospitals would have enough cases to be evaluated fairly and had enough an incentive to work on improving their performance. The conditions that

were used in the first two years of the program (fiscal years (FY) 2013 and 2014) included acute myocardial infarction (AMI), heart failure (HF), and pneumonia (PN). In FY 2015, the program added chronic obstructive pulmonary disease (COPD), and elective total hip arthroplasty and total knee arthroplasty. Depending on their excess readmission ratios for each of these included conditions, hospitals either avoided or received a penalty.

In fiscal year 2015, 78% of participating hospitals received a penalty, with an average penalty of a 0.63% decrease in base reimbursement amounts. The estimated total value of the penalties levied in fiscal year 2015 was $428 million, approximately a $199 million increase from the estimated value of the fiscal year 2014 penalties (Boccuti, 2015). This large increase in penalties is attributed to the greater number of diagnoses measured in the program as well as the changes to the maximum allowable penalty over the first three years of the program (Boccuti, 2015). With more conditions included, there was more potential for hospitals to have excess readmission ratios that hurt their overall program performance.

Excess readmission ratios (ERR) in each of the included conditions determine if a hospital will receive a reduction in their Medicare payments. Ideal hospital program performance requires having excess readmission ratios at or below 1 for all conditions. For this to happen, a hospital would have to perform better than the expected performance of an “average” hospital with the same case-mix, in every single included condition in the program. If a hospital has excess readmission ratios that are less than or equal to one for all of the included conditions, then that hospital will avoid a penalty. On the other hand, if a hospital has one or more conditions where their excess readmission ratio is above one, then they will receive a penalty in the applicable fiscal year manifested as a reduction in their base operating DRG payments by a percentage. The incentive structure has been in place for four years, beginning with a maximum

allowable penalty of 1% in fiscal year 2013 and a one percentage point increase in this maximum up to 3% in fiscal year 2015, where it will remain.

If a hospital has one or more excess readmission ratios above 1, they will receive a penalty. The actual size of that penalty depends on the total DRG reimbursements attributable to excess readmissions relative to the hospital's total DRG reimbursements during the measurement period. This proportion depends on how far above 1 the condition-specific ERR's are, the total DRG payments made for each specific condition, and the total DRG payments made for all discharges in that hospital over the same time frame. Given that the program penalizes hospitals based on their performance in a specific set of conditions, penalizes future Medicare reimbursements only, and calculates penalty size depending on the DRG volume of the pertinent conditions, there is inevitably a great deal of heterogeneity in the incentives that hospitals face. If a hospital has a low DRG volume of the conditions and performs poorly in only one condition, their incentives to make changes to their care trajectories should vastly differ compared to a hospital with a high DRG volume of the included conditions and poor performance in all five conditions. If two hospitals have the same readmissions performance in the program, but one hospital only relies on Medicare reimbursements for a very small share of their patients, while the other hospital relies on these reimbursements for a much larger proportion of their patients, do these hospitals try to improve to the same degree? Although the goal of the program is to reduce excess readmissions among Medicare patients, these aspects can drastically alter the actual incentives hospitals face to improve their performance. Accordingly, we are interested in exploring how much these sources of heterogeneity in hospital incentives are associated with hospital responses to the Hospital Readmissions Reduction Program.

Thus far, a lot of the literature surrounding the results of the Hospital Readmissions Reduction Program has focused on the issue of equity (Abelson, 2013). There is some evidence suggesting that the program has penalized hospitals serving low-income patients (Boccuti, 2015; Joynt and Jha, 2013). As a result, researchers and politicians have made calls for refinements to the program (Boozary et al., 2015). These findings have spurred a greater push to include socioeconomic status and community characteristics in the risk-adjustment process, as hospitals serving low-income patients may have a different set of barriers to overcome when trying to improve their readmissions performance (Boccuti, 2015). As part of this trend, the Institute of Medicine (IOM) has convened a committee to develop recommendations to Congress surrounding the inclusion of a social risk factor in Medicare payment programs (Keefe, 2016). As some authors have pointed out, the program's main purpose is to reduce excess readmissions, and there is evidence that national readmission rates began to fall in 2012 (Boccuti, 2015; Ness, 2013). While a few researchers have looked at the programmatic effects of the HRRP on readmission rates, these analyses have been at a fairly broad level, looking at average penalties by hospital characteristics. Analyses have found reductions in both all-cause readmissions as well as the condition-specific readmission rates over time, suggesting that the program may be having its intended effect. A recent study found a drop in the rate of readmissions that aligned with the introduction of the program, and did not see a significant increase in observational stays during the same time period. The authors concluded that the trends are consistent with the notion that hospitals are responding to the incentives in the program (Zuckerman et al., 2016).

We find that hospital performance in the program is strongly correlated with improvements over time, as hospitals with high excess readmission ratios in previous fiscal years had significantly larger changes in their performance compared to hospitals with lower excess

readmission ratios in prior years. Further, hospitals are improving their readmissions performance in many conditions where lower readmission rates reap higher marginal benefits on overall program performance as well as condition-specific performance. Hospital improvements in the conditions that were only added in FY 2015 were also correlated with the HRRP condition share of overall DRG reimbursements, while the share of inpatient days accounted for by Medicare patients was not related to improvements in excess readmission ratios across all five conditions.

Background of Hospital Readmissions Reduction Program

This nationwide focus on reducing readmissions has existed before the passage of the Affordable Care Act (ACA) in 2010 (Berenson et al., 2012). The Hospital Inpatient Quality Reporting Program (IQR) began requiring hospitals to report their 30-day readmission rates in 2009 as a follow-up to the reporting of hospital performance on patient experience surveys, mortality rates, and process of care measures. In the IQR program, hospitals are required to report their performance on included program measures to avoid reductions in their annual Medicare payment updates. The only requirement to avoid a reduction is that hospitals report their performance. Hospital performance in the IQR program is reported publicly on the website *Hospital Compare*, as their risk-standardized readmission rates were available starting in July 2009 (Dorsey, 2015). Originally, these rates were reported as being either above, below, or similar to the national average. This program increased the visibility of differences between hospitals in readmissions. The Hospital Readmissions Reduction Program was one of many programs that were legislated by the Affordable Care Act (ACA), established under section 3025 of the ACA. Similar to the other two programs that incentivize hospitals to improve their quality

of care (Hospital Value-Based Purchasing Program and Hospital-Acquired Condition Reduction Program), performance in the Hospital Readmissions Reduction Program determines if hospitals will receive a percentage point decrease in their base operating diagnosis related group (DRG) amount for a given fiscal year (Federal Register, 2012).

Participating hospitals in the program are acute care hospitals enrolled in the Medicare Inpatient Prospective Payment System. FY 2013 was the first year of the program that hospitals could receive a penalty based on their readmission performance. In HRRP, hospital performance is based on claims across three years. This extended data collection period is used to ensure that enough cases (minimum of 25) are collected to reliably compare hospitals to one another. Table 2.1 outlines the time periods of the program for the first four years.

In FY 2016, data were collected from participating hospitals from July 1st, 2011 through June 30th, 2014. This allows enough time for CMS to process data, calculate the excess readmission ratio, and apply any penalties to hospitals before FY 2016 begins. Each year of the program, the maximum penalty percentage has increased by 1 percentage point, plateauing at 3% from FY 2015 onwards.

Table 2.1 - Program Properties

	FY 2013	**FY 2014**	**FY 2015**	**FY 2016**
Readmissions Timeline	7/1/08-6/30/11	7/1/09-6/30/12	7/1/10-6/30/13	7/1/11-6/30/14
Maximum Penalty	1.0%	2.0%	3.0%	3.0%
Included Conditions	Pneumonia, Heart Attack, Heart Failure	Pneumonia, Heart Attack, Heart Failure	Pneumonia, Heart Attack, Heart Failure, COPD, Hip or Knee Replacement	Pneumonia, Heart Attack, Heart Failure, COPD, Hip or Knee Replacement

Penalties for hospital *i* in the program are determined using the following equation, where the HRRP Adjustment Factor is the official name of the penalty:

$$HRRP\ Readmissions\ Adjustment\ Factor_i = 1 - \left(\frac{Aggregate\ Payments\ for\ Excess\ Readmissions_i}{Aggregate\ Payments\ for\ All\ Discharges_i}\right)$$

(Equation 1)

If the proportion of payments made for excess readmissions to payments made for all discharges is greater than the maximum allowable penalty in that fiscal year (e.g., 3% in FY 2015), then the hospital receives the maximum penalty. To translate this rule within the construct of Equation 1, if the *HRRP Adjustment Factor* is smaller than *1 – Maximum Allowable Penalty*, then a hospital receives the maximum penalty. The maximum penalty changes all values less than *1 – Maximum Allowable Penalty* to the maximum penalty. Then, in the applicable fiscal year, this *HRRP Adjustment Factor* is multiplied by the base-operating DRG amount for a hospital; if a hospital avoids a penalty, then the adjustment factor is 1, otherwise, it is a value lower than 1 and a hospital would receive a smaller base-operating DRG amount. The *aggregate payments for all discharges* is the sum of base-operating DRG payments for all discharges in a hospital during the performance period. *Aggregate payments for excess readmissions* are calculated using the following equation for hospital *i* and measured condition *j*:

$$Aggregate\ Payments\ for\ Excess\ Readmissions_i = \sum_{j=1}^{J} Base\ Operating\ DRG\ Payments_{i,j} * \left(Excess\ Readmission\ Ratio_{i,j} - 1\right)$$

(Equation 2)

For this part of the formula, CMS needs to estimate the amount of condition-specific payments that went towards excess readmissions. So CMS multiplies the total base operating DRG payments made for each condition in each hospital by the share of readmissions for the

condition that were considered to be excess. Summing this amount across all conditions gives the total DRG payments made for excess readmissions. The condition-specific base operating DRG payments are collected from the claims data and compiled by CMS. The other part of this calculation is the excess readmission ratio. This value is calculated for each condition *j* in hospital *i* as:

$$Excess\ Readmission\ Ratio_{i,j} = \frac{Risk\ Adjusted\ Predicted\ Readmissions_{i,j}}{Risk\ Adjusted\ Expected\ Readmissions_{i,j}}$$

(Equation 3)

The excess readmission ratio is calculated by dividing the risk-adjusted predicted number of readmissions by the risk-adjusted expected number of readmissions. Broadly, the methodology CMS uses compares a hospital's readmissions performance given its case-mix with the expected performance of an average hospital with the same case-mix. The hospital case-mix relies on patient gender, age, and condition-specific clinical risk factors. Using a hierarchical random effects logistic regression model, CMS calculates the expected and predicted number of readmission for each hospital by condition. For the numerator, CMS sums the probabilities of readmission within 30-days for all included patients in a hospital for a given condition. For each patient *p* included as an eligible diagnosis for a hospital *i*, the probability of a readmission in the numerator depends on the following logistic specification:

$$h(Y_{ip}) = \alpha_i + \beta Z_{ip} \text{ where } \alpha_i = \mu + \omega_i \text{ and } \omega_i \sim N(0, \tau^2)$$

so

$$h(Y_{ip}) = \mu + \omega_i + \beta Z_{ip}$$

In the equation above, μ is the "adjusted average-hospital effect" across all hospitals, and τ^2 is the between-hospital variance, and α_i is the hospital-specific effect. Z_{ip} is the set of patient level

risk factors (age, gender, and condition-specific clinical factors). The predicted probability for each patient in the numerator is:

$$e^{(\mu+\omega_i+\beta Z_{ip})} \Big/ (1+e^{(\mu+\omega_i+\beta Z_{ip})})$$

Then, CMS sums the predicted probability across all *N* eligible patients at hospital *i*:

$$\sum_{n=1}^{N} e^{(\mu+\omega_i+\beta Z_{ip})} \Big/ (1+e^{(\mu+\omega_i+\beta Z_{ip})})$$

This predicted number of readmissions is then divided by the number of eligible patients to provide a predicted readmissions rate. So for every hospital, their observed effect on readmission likelihood is captured through their difference from the average hospital effect on readmissions. For the denominator, CMS estimates the probability of a readmission for a hospital's case-mix given the performance of an average hospital. They remove the hospital-specific effect, so the specification is:

$$h(Y_{ip}) = \mu + \beta Z_{ip}$$

and again, the predicted probability for an individual patient becomes:

$$e^{(\mu+\beta Z_{ip})} \Big/ (1+e^{(\mu+\beta Z_{ip})})$$

CMS sums this predicted probability over all eligible patients *N* for that hospital within that condition to come up with an expected number of readmissions:

$$\sum_{n=1}^{N} e^{(\mu + \beta Z_{ip})} \Big/ (1 + e^{(\mu + \beta Z_{ip})})$$

As is done with the numerator, this expected total of readmissions is divided by the total number of eligible patients to provide an expected readmissions rate. The excess readmission ratio that is reported on CMS' *Hospital Compare* website is the ratio of the predicted readmission rate to the expected readmission rate.

Index admissions for Medicare fee-for-service patients who are aged 65 or over, are discharged alive and not transferred to another acute care facility, and with a principal discharge diagnosis matching one of the necessary conditions are included in this program. Within 30 days of discharge, all unplanned readmissions are counted for hospitals by condition. Although the definition of excluded readmissions has changed over the years, in general, planned readmissions, same-day readmissions, observation stays, emergency department visits, and admissions to non-short-term acute care hospitals are not included as readmissions in the measures (Dorsey, 2015). A significant change in the program occurred in FY 2014, when CMS introduced an algorithm to account for planned readmissions starting in FY 2014 of the program. Finally, even if a patient has multiple unplanned readmissions, the measure uses a simple binary variable to indicate whether or not a patient with a given condition has had an unplanned readmission within the 30-day timeframe (McIlvennan et al., 2015; Centers for Medicare & Medicaid Services, 2013). More details are provided by CMS on the QualityNet website (Dorsey, 2015).

Hospital A

Condition	ERR	DRG Sum	DRG Weights for Excess Readmissions
AMI	1.2	338	67.6
HF	1.4	450	180
PN	1.3	353	105.9
HK	1.2	548	109.6
COPD	1.2	315	63
Total Excess DRG			526.1
Total DRG	14726		
Adjustment Factor = 0.964274			
Penalty = 3%			

Hospital B

Condition	ERR	DRG Sum	DRG Weights for Excess Readmissions
AMI	0.99	338	0
HF	1.1	450	45
PN	1.1	353	35.3
HK	1.25	548	137
COPD	1.25	315	78.75
Total Excess DRG			296.05
Total DRG	14726		
Adjustment Factor = 0.979896			
Penalty = 2.01040%			

Hospital C

Condition	ERR	DRG Sum	DRG Weights for Excess Readmissions
AMI	0.99	338	0
HF	0.98	450	0
PN	0.97	353	0
HK	0.96	548	0
COPD	1.05	315	15.75
Total Excess DRG		15.75	
Total DRG	14726		
Adjustment Factor = 0.99893			
Penalty = 0.10700%			

Figure 2.1 - Sample HRRP Calculation

Figure 2.1 illustrates the methodology used to calculate hospital performance in this program. In this example, Hospital A has excess readmission ratios that are above 1 for all five conditions in the program, Hospital B has one condition with an excess readmission ratio below 1, while all the others are above 1, and Hospital C has only one condition with an excess readmission ratio above one, while the remaining four are below one. In this example, all three hospitals have the same condition-specific DRG amounts, as well as the same total DRG amount of 14,726. The condition-specific DRG amounts are sums of the transfer-adjusted DRG relative weights for each condition across the relevant data time period. The total DRG amount is the sum of all transfer-adjusted DRG relative weights for all conditions in the data time period. The DRG amounts used in this example are the means of these parameters from our analytic sample.

The only difference between each of these hospitals in terms of the program performance calculation in this example is in their ERR's. This difference results in varying values of the numerator in equation 1, which leads to differences in their calculated *Adjustment Factors*. To replicate the methodology CMS uses, the difference between the *ERR's* and 1 is multiplied by the *DRG Sum* for each condition, resulting in the *DRG Payments for Excess Readmissions.* Only if the ERR for a condition is greater than 1 does the DRG payment get added with the DRG amounts from the other conditions with excess readmissions. The sum of this value across all the conditions is the numerator in equation 1, while the *Total DRG* is the denominator. As seen above, differences in excess readmission ratios across the five program conditions can result in quite different Adjustment Factors, when holding all other variables constant.

Theoretical Framework and Hypotheses

Readmissions are on average decreasing in the United States, after remaining relatively stagnant from 2004-2009 (Gerhardt et al., 2013; McIlvennan et al., 2015; Zuckerman et al.,

2016). We are interested here in identifying what may be important drivers of program performance over time. Using changes in condition-specific, hospital-level excess readmission ratios over multiple fiscal years of the program, we are interested in understanding if the heterogeneous incentive structure is associated with differences in hospital performance. We focus our analysis on four characteristics that play a role in how hospitals interact with the incentive structure of the HRRP and the methodology CMS uses to determine program performance:

1) Medicare inpatient proportion
2) Program condition DRG Weight
3) Marginal benefits of improving performance in each condition
4) Previous year program performance

Medicare Inpatient Proportion

To maintain simplicity and uniformity, the program applies the same incentive structure for hospitals, as long as they meet the minimum case requirement. Hospitals with Medicare patients accounting for as low as only 10% of their inpatient days face the same penalty percentages as hospitals with a much higher share of Medicare-financed inpatient days. The proportion of inpatient days accounted for by Medicare patients varies widely (see Figure 2.2). Since the financial penalty in the program only applies to future Medicare reimbursements, the volume of reimbursements at stake depends on the share of Medicare patients a hospital sees relative to patients from other insurance sources. Presumably, if a hospital has a much larger proportion of its revenue generated from privately insured patients, the risk of a percentage point reduction in their Medicare reimbursements may be less of a driving force for change.

Conversely, for hospitals where the majority of inpatient days are reimbursed by Medicare, the incentives may be large enough to spur improvements in readmissions. Other research has shown that high Medicare or Medicaid populations may be associated with slower adoption of technologies to improve care (Menachemi et al., 2007). Therefore, the Medicare inpatient proportion is potentially an important driver of a hospital's response to incentives in the Hospital Readmissions Reduction Program.

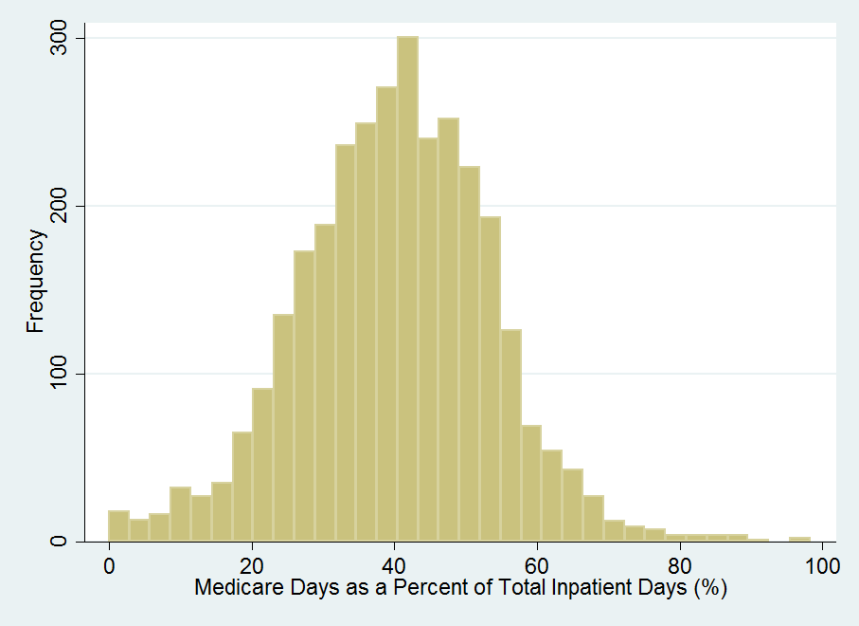


Figure 2.2 - Distribution of Medicare Proportion of Inpatient Days

For hospitals where Medicare patients make up a large portion of the patient care population, the financial incentive they face to avoid a reduction in their reimbursements should be greater than hospitals where Medicare patients account for a small portion of the patient population. If financial incentives are the main driving force behind hospital behavior, then a hospital that predominantly relies on non-Medicare reimbursements would have a smaller incentive to improve their program performance.

Hypothesis 1: Hospital Medicare patient proportion of total inpatient days will be positively correlated with improvements in excess readmission ratios across all conditions.

Program Condition DRG Weight

As can be seen in equations 1 and 2, the excess readmissions reimbursement amount is a vital determinant of overall program performance. CMS compares this amount to the sum total of each hospital's overall DRG reimbursements (Equation 1). This ratio is driven in part by the relative volume of the program conditions within each hospital. While CMS chose the conditions to be included in the program due to their high-volume characteristic at a national level, there is heterogeneity in condition-volume at the hospital level. Since the program incentivizes multiple conditions, and CMS calculates hospital performance using the total DRG payments across all included conditions, hospitals may be influenced by the overall volume share of the program-specific conditions. If two hospitals had the same exact excess readmission ratios across all five conditions, but for one hospital the DRG share of the program conditions was much higher than for the other hospital, then the output of equation 2 would be different. Figure 2.3 shows the total DRG weight of all five program conditions divided by the overall DRG reimbursements for each hospital in FY 2015 of the program. For some hospitals, the HRRP conditions account for less than 10% of their overall DRG reimbursements, while for other hospitals, this proportion is greater than 25%. As hospitals face resource constraints and thus cannot invest in all quality improvement efforts, this type of variation may be important. Figure 2.4 shows the differences that can exist in hospital program performance when the only source of variation is the DRG weight of the 5 program conditions. In this example, Hospital A and D have the same ERR's across all 5 conditions as well as the same overall DRG reimbursement amount, while their

condition-specific DRG's are different. Hospital D has smaller condition-specific DRG totals than Hospital A.

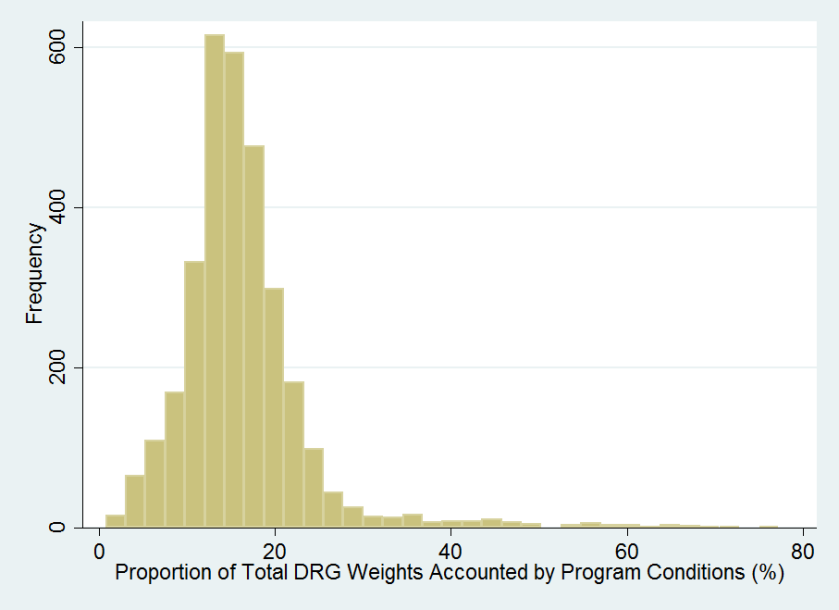


Figure 2.3 - Program Condition DRG Weights

Hospital A

Condition	ERR	DRG Sum	DRG Payments for Excess Readmissions
AMI	1.05	338	16.9
HF	1.1	450	45
PN	1.2	353	70.6
HK	1.02	548	10.96
COPD	1.04	315	12.6
Total Excess DRG			156.06
Total DRG	14726		
Adjustment Factor = 0.989402			
Penalty = 1.0598%			

Hospital D

Condition	ERR	DRG Sum	DRG Payments for Excess Readmissions
AMI	1.05	238	11.9
HF	1.1	350	35
PN	1.2	253	50.6
HK	1.02	448	8.96
COPD	1.04	215	8.6
Total Excess DRG			115.06
Total DRG	14726		
Adjustment Factor = 0.99219			
Penalty = 0.7810%			

Figure 2.4 - Sample HRRP Calculation 2

Accordingly, the result of equation 2 and the numerator in equation 1 is smaller for Hospital D compared to Hospital A, leading to a smaller penalty for Hospital D even though the ERR's were

the exact same across the hospitals. As the proportion of the total program-specific condition DRG sum to overall DRG reimbursement amounts rises, the incentives for hospitals to improve their excess readmission ratios should be greater. Holding all else constant, a higher proportion results in a larger penalty. Therefore, the incentive to improve will be higher as this proportion rises.

Hypothesis 2: The proportion of overall DRG reimbursements accounted by the program conditions will be positively correlated with hospital program improvement over time.

Marginal benefits of improving performance in each condition

Improvements in condition-specific readmissions do not have a uniform effect on overall program performance. The CMS methodology incorporates both condition-specific DRG volume and the ratio of risk-adjusted predicted to risk-adjusted expected readmissions. Therefore, it is not necessarily true that a reduction in predicted readmissions probability for a patient with an AMI would have the same effect on program performance as the same reduction in the predicted readmission probability for a patient with heart failure or pneumonia. Accordingly, each hospital faces different marginal benefits of their efforts to reduce readmissions in the program conditions. If hospitals want minimize penalties in this program, they would work to improve readmissions in the conditions that have the highest potential to improve their HRRP Adjustment Factor. We identify this marginal improvement as a decile improvement in the difference between the risk-adjusted predicted and expected readmission rates, allowing us to standardize improvements in performance regardless of the number of eligible patients a hospital may have.

With three fiscal years of experience with the program, and since hospitals receive reports detailing their performance in all conditions compared to other hospitals, participating

hospitals are now likely able to identify which condition improvements result in the greatest effects on program performance.

Hypothesis 3: Hospitals with a larger marginal effect of improvements in performance will have larger improvements than those with smaller marginal effects.

Previous year program performance

The Hospital Readmissions Reduction Program stems from efforts to improve hospital quality while using public reporting. Every year of the program, information on quality is made publicly available so that anyone can go to CMS' website and look at a hospital's performance. Given the visibility of results, hospitals may be motivated to improve their readmissions to avoid bad publicity in future years. Accordingly, we believe there may be two important motivating factors that drive hospital response in this program – a) achieving a symbol of "quality" by being a hospital that is not penalized, and b) improving their excess readmission ratio. For hospitals that place value on attaining a symbol of quality, they might be incentivized to bring their excess readmission ratio below 1 for the included conditions. Similarly, poor performance in a previous fiscal year of the program could drive hospitals to improve their excess readmission ratio, irrespective of penalty avoidance. Because there are two incentivized "goals" in this program (above/below 1, reducing excess readmission ratio), there is a non-linearity in the incentive structure. We can exploit this non-linearity based on how we characterize program performance in previous years.

Hospitals that receive a penalty are incentivized to improve their performance on readmissions to avoid penalties in future years. The underlying assumption of this type of

incentive program is that levying penalties to hospitals will spur improvements in performance. Accordingly, hospitals that received a penalty in previous years should have a significantly higher drop in their excess readmission ratios compared to hospitals that avoided a penalty. Fiscal year 2014 results are released after the performance period for fiscal year 2015 has ended, but since performance in this program is a 3-year moving average, hospitals should know approximately how well they are performing.

Hypothesis 4a: Hospitals that receive a penalty are expected to make significant improvements in their readmissions performance over the fiscal years of our analysis.

The highest 25th percentile of excess readmission ratios above 1 will be the set of hospitals with the worst performance in each condition for a fiscal year. The room for improvement for these hospitals is much greater than all other hospitals, so it is expected that these hospitals will improve their performance over time. Since the hospitals at the tail end of the excess readmissions ratio have the most room for improvement, even small changes should result in improvements in penalty size in subsequent years.

Hypothesis 4b: Hospitals with the worst excess readmission ratios above 1 are expected to make larger improvements to their excess readmission ratios compared to hospitals that have excess readmission ratios right above 1. Similarly, increased excess readmission ratios in previous years are expected to be significantly associated with improvements over time.

Condition-Specific Responses

The lag between performance periods and data release prevents hospitals from making changes that would improve their scoring in the following year of the program. For certain conditions though, hospitals should know where they stand compared to others and how much they can improve based on previous fiscal year data. For the conditions added in FY 2015, the hospitals do not have this information. We would then expect that in FY 2015, the correlation between poor previous performance and improvements over time to be high for the original three conditions, and low for chronic obstructive pulmonary disease (COPD) and elective total hip arthroplasty and total knee arthroplasty.

With the program in place for multiple fiscal years, hospitals should have a good sense of their performance in the original three conditions, and since program performance is based approximately on a 3-year moving average, they should be able to use that knowledge to identify whether they need to make large improvements or not. For the conditions that were added only in FY 2015, this learning is less likely to have happened already, so poor performance in the first year may have less of a correlation with changes over time.

Hypothesis 5: Poor performance, as captured by the raw excess readmission ratio, in fiscal year 2015 is expected to be associated with greater improvements for AMI, Pneumonia, and Heart Failure, but not for COPD and total hip or knee arthroplasty.

Data

CMS releases publicly available data for all years of the program on *Hospital Compare* with hospital-level data detailing the excess readmission ratios for each condition, the number of

eligible cases for each condition, and the overall Readmissions Adjustment Factor. Since CMS added an algorithm to exclude planned readmissions in FY 2014, we restrict our analysis to FY 2014 – FY 2016 to maintain a degree of uniformity in the calculation of excess readmission ratios. In FY 2015, CMS began providing detailed data regarding the number of cases and the DRG weights by condition for each year in the performance period. Data at this level of detail is necessary to translate excess readmission ratios into the overall Hospital Readmissions Reduction Adjustment Factor. The CMS *Impact File* is released every year, and provides data on the Adjustment Factor, as well as the proportion of inpatient days accounted for by Medicare patients. The American Hospital Association (AHA) Annual Survey data allows us to control hospital performance by certain hospital characteristics, including bed size, teaching status, ownership, and region. We restrict our analyses to hospitals that have all pertinent data across all years of the analysis, resulting in 3,125 hospitals. For our analyses, we require data on HRRP performance across the three fiscal years, condition-specific DRG weights, and the predicted and expected readmission rates. 188 hospitals did not have program performance data for both FY 2015 and FY 2016, and the remaining hospitals that were dropped did not have complete DRG data or the predicted and expected readmission rates available. The overall hospital sample size is 3,125, but the samples within the analyses for each condition differ, depending on whether or not hospitals had the required 25 eligible cases. Our analytic sample includes 90% of the hospitals that participated in the program in FY 2016.

Methods

The analytic framework uses linear regression to identify whether sources for heterogeneity in the incentives hospitals face in the Hospital Readmissions Reduction Program are correlated with changes in excess readmission ratios over time. We analyzed changes in

excess readmission ratios (ERR) within each condition in the program. For excess readmission ratios, higher values indicate worse performance, since this means the risk-adjusted predicted readmission rates were higher than the risk-adjusted expected rates. We specify the outcome variable as $ERR_t - ERR_{t-1}$ or $ERR_t - ERR_{t-2,}$ so a negative value indicates improvement.

To calculate the proportion of DRG reimbursements accounted for by the five HRRP conditions in FY 2015, we calculate the sum of the transfer-adjusted condition-specific DRG amounts for each of the fiscal years in the performance period. This sum is then the total transfer-adjusted DRG weights for all program conditions across the whole performance period. Dividing this amount by the sum of all transfer-adjusted DRG weights in the performance period provides us with DRG proportion accounted for by the HRRP conditions. We then generate a binary indicator variable for hospital performance in previous years (FY 2014 and FY 2015), that is equal to 1 if a hospital has a HRRP Adjustment Factor less than 1, and 0 when the Adjustment Factor is equal to 1. It is also possible that there are nonlinear relationships between some of these variables and hospital responses in the program. Hospitals with very large proportions of Medicare patients may be incentivized to improve because of the volume of patients, while hospitals with a small proportion may be incentivized to improve because of the few patients responsible for their penalties. Accordingly, we also specify the Medicare proportion and DRG proportion variables as categorical variables using the quartiles of the distribution to examine any potential non-linear effects.

To generate condition-specific performance indicators, we use both the raw excess readmission ratio values from previous years, but also have specifications where we use indicator variables for an ERR being above 1 or the quartile of an ERR when above 1. Since there is a non-linearity in the incentives hospitals face with regard to the ERR (ERR≤1: condition data not

included in penalty calculation, ERR>1: included in penalty calculation), we also explore specifications where the previous ERR's are estimated using a piecewise approach. For this, we use a spline with a knot at 1. These alternative methods of capturing previous year program performance allow us to explore the many ways in which previous performance may or may not be correlated with improvements. For each fiscal year in the program, program performance results are released almost a year after the performance period ends (FY 2014 performance period was 7/1/09-6/30/12, and FY 2014 Final Rule was published in 8/2013). Thus, hospitals may not know their excess readmission ratios as well as overall program performance until after the performance period for the following year is completed, even with the hospital preview period before data is made publicly available. The hospital preview period usually occurs from mid-June to mid-July, a month before the final rule is published in August. Accordingly, program performance from two fiscal years ago may be more useful in driving hospital behavior.

Marginal Effect Calculation

As the equation for the ERR shows, the difference between the numerator and denominator is driven primarily by the ω_i value, the hospital-specific effect on readmissions probability. Since CMS sums probabilities, the actual difference between the risk-adjusted expected and predicted number of readmissions is driven by the sum of these hospital-specific effects over all eligible cases. Yet, since the model is not an ordinary least squares specification, the transformation required to generate predicted probabilities and the subsequent summation of these predicted probabilities over all eligible cases prevents us from specifying the difference between the risk-adjusted predicted and expected readmissions in an easily interpretable form. Intuitively, though, this difference is driven primarily by the ω_i value for each hospital within

each condition. Taking the numerical difference between the predicted and observed rates provides us with an approximation of the variation in hospital-specific effects. The distribution of these values for pneumonia in fiscal year 2015 match what we would expect (Figure 2.5). While we cannot accurately estimate ω_i without having all the claims data CMS uses to generate the ERR's, we believe the differences between the risk-adjusted predicted and expected rates is a reasonable proxy for our purposes. In order for a hospital with an ERR above 1 to decrease the difference between the predicted and expected number of readmissions, they must improve their performance on readmissions so that their calculated hospital-specific effect gets closer to, or smaller than the average hospital effect ($\omega_i\ decreases$). Accordingly, the marginal benefit calculation uses a standardized improvement in the predicted readmission rate for each hospital within each condition to determine the effect on condition-specific as well as overall program performance. To improve the predicted readmission rate, a hospital would have to reduce readmissions enough to decrease their hospital-specific effect. Since a one readmission reduction would have varying effects on the hospital-specific effect based on condition volume, this approach allows us to standardize the improvement across hospitals.

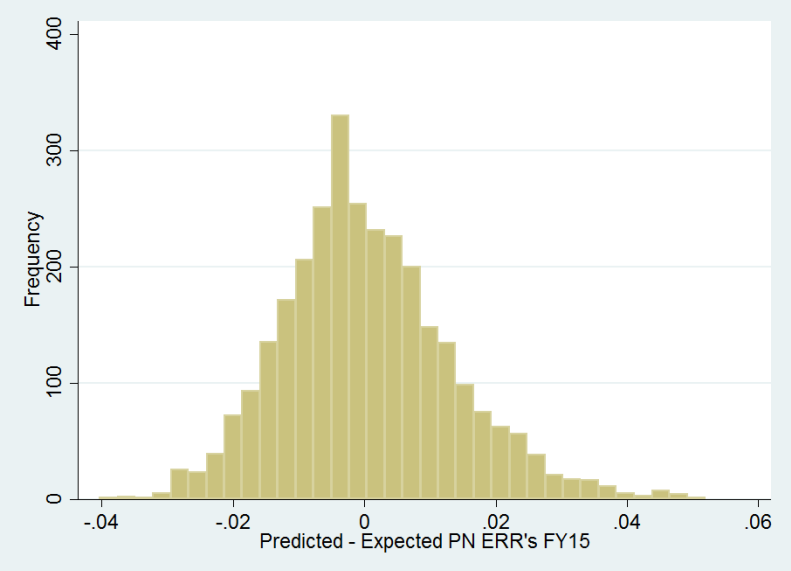


Figure 2.5 - Raw Difference in Predicted and Expected Rates

We calculate the effect of a decile improvement for each hospital within each condition, in this calculated difference. While the effect on the ERR of a one readmission decrease is very tangible, the methodology CMS uses prevents this calculation without detailed claims data. Furthermore, a one readmission decrease would only be captured as an extremely miniscule change in ω_i, and would be driven to a degree by the volume of eligible patients for each condition (one less readmission among 25 patients may not be equal to one less readmission among 1000 patients). Instead, to standardize improvements in performance, we believe a decile improvement in the difference in the risk-adjusted predicted and expected rates provides a tangible way to characterize a marginal improvement in hospital performance. A one decile improvement in this difference should capture a marginal improvement in condition-specific readmissions performance, while minimizing sample size effects.

$$\begin{aligned} &Estimated\ Predicted\ Readmission\ Rate_{ij\ ,2010-2013} \\ &\quad = Excess\ Readmission\ Ratio_{ij,2010-2013} \\ &\quad * Risk\ Adjusted\ Expected\ Readmission\ Rate_{ij,2010-2013} \end{aligned}$$

With this estimated predicted readmission rate, we then calculate the difference between the predicted and expected rates for each hospital *i* in condition *j*.

$$\begin{aligned} &Estimated\ Difference_{ij\ ,2010-2013} \\ &\quad = Estimated\ Predicted\ Readmission\ Rate_{ij,2010-2013}\ ) \\ &\quad - Expected\ Readmission\ Rate_{ij,2010-2013} \end{aligned}$$

Then, we improve this estimated difference by a decile within the distribution, recalculate the new predicted readmission rate and divide by the expected readmission rate to provide a new estimate of the Excess Readmission Ratio conditional on this marginal improvement:

$$Excess\ Readmission\ Ratio'_{ij,2010-2013} = \frac{Predicted\ Readmission\ Rate'_{ij,2010-2013}}{Risk\ Adjusted\ Expected\ Readmission\ Rate_{ij,2010-2013}}$$

Plugging this new excess readmission ratio at the condition specific level into the necessary formulas leads to:

$$\partial(HRRP\ Adustment\ Factor_{i,2010-2013}) \Big/ \partial(Predicted\ Readmission\ Rate_{ij,2010-2013})$$

Multiplying the condition-specific excess readmission ratio by the risk-adjusted expected readmission rate we generate an estimate for the risk-adjusted predicted readmission rate. We use this value instead of the provided risk-adjusted predicted readmission rate in the data releases because there are some data alignment errors in the CMS provided files when comparing the actual ERR with the result of dividing the predicted rate by the expected rate. Thus, to better align the released data, we use the risk-adjusted expected rate and the actual excess readmission ratio to estimate the risk-adjusted predicted readmission rate. Taking the difference between the provided expected readmissions rate and the calculated predicted readmissions rate results in our proxy for the hospital-specific effect. Decreasing this difference by a one decile improvement in the distribution provides us with the necessary data to recalculate a new predicted readmission rate conditional on hospitals improving their performance enough to move their hospital-specific effect. A one decile improvement is calculated by taking each predicted minus expected value, and changing it to the predicted minus expected difference of a hospital that is one decile better in this value. For example, a hospital with a very poor excess readmission ratio will have a positive value when subtracting the expected readmission rate from the predicted readmission rate. When improving this value by a decile, the difference between the predicted and expected readmission rates will decrease in absolute terms, and may even become negative. Subsequently, this "new" difference will result in an estimated lower excess readmission ratio conditional on a marginal decrease in readmissions.

Calculating the effect of a marginal improvement in readmissions for condition *j* on the overall Readmissions Adjustment Factor (which takes into account the excess readmission ratios for all conditions and their base operating DRG payments for each condition), provides estimates of the effect of changing readmissions for each condition on the program adjustment factor. We are accurately able to replicate the actual Adjustment Factor in fiscal year 2015 for 62% of hospitals in our analytic sample, while for the remaining 38%, the discrepancies between our calculated adjustment factor and the CMS provided numbers are between -0.0025 and 0.0015. For the 62% of hospitals kept in this sub-sample, we are able to exactly match their Adjustment Factor in FY 2015. Regressing actual changes in readmission ratios on marginal benefits of improvement, we are able to determine if hospitals are making changes in the conditions with the greatest impact on their performance in the program. Since the fiscal year 2014 results come out in time to hypothetically impact the FY 2016 results, we also calculate the marginal benefit of performance improvement on changes in the ERR. Unfortunately, CMS did not release all the necessary data to calculate the HRRP Adjustment Factor until FY 2015. Instead, we calculate the change in condition-specific excess readmission ratios for a marginal improvement in the hospital-specific effect.

The change in ERR's (calculated as $ERR_{marginal} - ERR_{original}$) are all negative, which gives us confidence that with our decile change in the predicted rate, we are improving hospital performance by leading to a smaller ERR (Figure 2.6 shows the distribution of changes in heart failure ERRs conditional on this decile improvement).

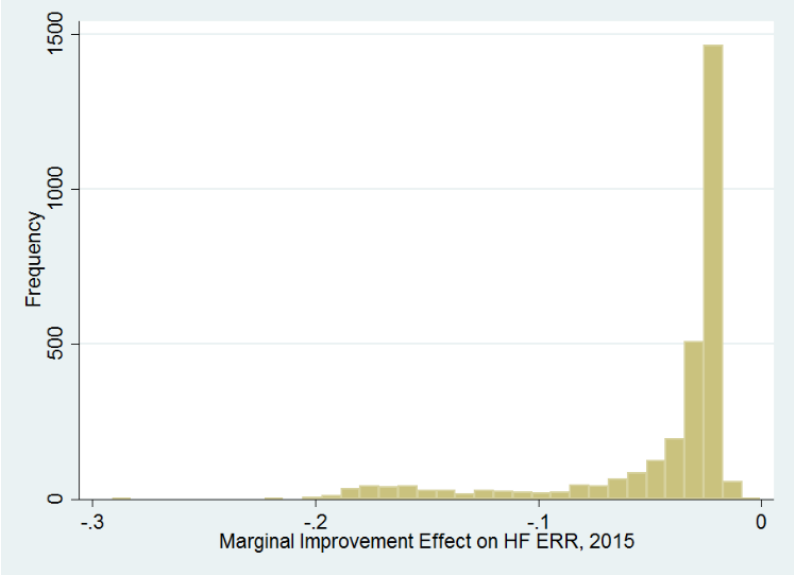


Figure 2.6 - Marginal Benefit of Standardized Improvement

Table 2.2 shows the average change in the FY 2015 Adjustment Factor for all five conditions if hospitals had improved their performance by a decile, the average change in ERR's for all five conditions in FY 2015 if hospitals had improved their performance by a decile, and the average change in the ERR for the three conditions in FY 2014 if hospitals had made marginal improvements in performance. The data for the Adjustment Factor changes is only shown for those hospitals where we were able to exactly replicate their Adjustment Factor, as the discrepancies in Adjustment Factor replication could skew our results if we did not exclude those hospitals.

Table 2.2 - Average Benefit of a Marginal Improvement on HRRP Adjustment Factor and Excess Readmission Ratio

	AMI	**HF**	**PN**	**HK**	**COPD**
Marginal Effect on Adjustment Factor (FY 15)	0.00025	0.00052	0.00052	0.00094	0.00041
Marginal Effect on ERR (FY 15)	-0.037	-0.043	-0.039	-0.071	-0.034
Marginal Effect on ERR (FY 14)	-0.043	-0.040	-0.040	-	-

Note: AMI – acute myocardial infarction, HF – heart failure, PN – pneumonia, HK – Hip or Knee Arthroplasty, COPD – chronic obstructive pulmonary disease

There are differences in the average marginal benefit of the decile improvement on ERR's as well as the overall program Adjustment Factor. On average, improvements in the hip or knee replacement condition have the largest effect on program performance. The hospitals included in the sample for the marginal effect on the FY 15 HRRP Adjustment Factor is limited to those hospitals that received a penalty in FY 15, as Adjustment Factors cannot go above 1, so the ceiling of marginal benefit is reached for those hospitals. These average effects of a marginal improvement are in the expected direction, as a marginal improvement in performance should lead to a reduction in excess readmission ratios, and an increase in the HRRP Adjustment Factor (reduction in penalty). Multiplying the changes in the predicted readmission rates by the condition-specific discharges, we calculate that the average readmissions reduction captured in our marginal improvement estimation varies from 2.0-4.3 across all 5 conditions in FY 2015, with the standard deviation ranging from 3-6.9.

Empirical Approach

To identify areas of heterogeneity in hospital response to the incentives of the Hospital Readmission Reduction Program, we estimate the following reduced form equation:

$$\Delta ERR_{i,j} = \alpha + \beta_1 MedicarePct_i + \beta_2(DRG\ Weight_i) + \beta_3(MarginalEffect_{i,j}) + \beta_4(Performance_{i,j}) + \beta_5 Z_i + \mu_{i,j}$$

In this specification, $\Delta ERR_{i,j}$ is the change in the excess readmission ratio for hospital *i* in condition *j* between FY 2016 and FY 2015 or FY 2014. Positive coefficients mean an increase in the independent variable was associated with a worsening of performance over time, while negative coefficients mean an increase in the independent variable was associated with an

improvement in readmissions performance. The $MedicarePct_i$ variable is the proportion of inpatient days accounted by Medicare patients. $DRG\ Weight$ is the sum of the five program-specific condition DRG reimbursements divided by the total DRG reimbursements for each hospital in FY 2015. $MarginalEffect$ is the estimated effect of the marginal improvement in the hospital-specific effect within each condition on the ERR or the Adjustment Factor. Finally, $Performance_{i,j}$ is the measure of hospital performance in condition j in the previous fiscal years, specified in different ways as mentioned earlier. Since FY 2014 results come out early enough to impact FY 2016 performance, we include these in the primary specification. Adding in a vector of hospital characteristics (Z_i) to this model allows us to capture the effects of fixed hospital characteristics on changes in readmission rates. Because the four factors identified here are not necessarily exogenous to hospital performance improvement, adding these potential confounding variables helps reduce the bias on our coefficients. The specific hospital characteristics we include in our specifications are bed size, hospital ownership, teaching status, and geographic location.

We build up to this overall specification by first running analyses where we regress the change in excess readmission ratios on each of our independent variables of interest, first in a bivariate specification, and then add the set of hospital controls. It is not apparent a priori that hospitals respond, if at all, to these sources of incentive heterogeneity in unison. These independent variables of interest could vary in importance based on an individual hospital's quality improvement context and characteristics. Accordingly, we first evaluate the bivariate relationships between these variables and changes in performance over time, and then add hospital-level controls. We end our analyses with the full specification of our empirical approach.

Hypothesis 1 related to the change in performance over time for hospitals based on their proportion of Medicare patients. Since we expect the incentive to improve to be greater for hospitals with a greater share of Medicare patients, our hypothesis is that an increase in $MedicarePct$ should be associated with a significant change in ERR's, and thus we expect $\beta_1$to be negative and significant. When looking at the relationship between changes in ERR's and the total DRG weight of the program conditions relative to the total DRG weight in a hospital, our expectation is that the higher the proportion of the overall DRG weights that the program conditions accounted for, the greater the incentive to improve. Thus, our hypothesis is that β_2 will be negative and significant, where increases in the program condition DRG proportion are associated with subsequent improvements in excess readmission ratios.

Turning to our marginal effect variable, after multiple years in the program, we expect hospitals to be able to identify which conditions would help improve their program performance the most. Thus, our hypothesis is that β_3 will be negative and significant, where higher marginal benefits of a decile improvement are associated with significant improvements in excess readmission ratios over time. In terms of performance in the program in the previous years, we expect that receiving a penalty will drive hospitals to improve (Hypothesis 4a), and that hospitals with very high excess readmission ratios will have a large incentive to improve (Hypothesis 4b). While we explore previous program performance with multiple specifications, we generally expect β_4 to be significant and negative (penalty receipt or higher ERR associated with greater improvement). Finally, since there are differences in which conditions hospitals have had to focus on, as well as the amount of time hospitals have had to respond and react to their program performance, we expect β_4 to be significant and negative for AMI, pneumonia, and heart failure, and insignificant for COPD and total hip/knee arthroplasty.

Results

Of the 3,125 hospitals in our analytic sample, there was little change in terms of avoiding or receiving a penalty across years, as 94% and 93% of hospitals that received a penalty in FY 2014 also received a penalty in FY 2015 or FY 2016 respectively. In each successive year of the program, the percentage of hospitals receiving a penalty increased. In the most recent year (FY 2016), 79% of participating hospitals received a penalty, up from 66% in FY 2014. This jump in the number of hospitals receiving a penalty has been attributed to the increased number of included conditions beginning in FY 2015. The majority of hospitals in the analytic sample are small (<200 beds), non-teaching hospitals, non-profit, and located in the southern United States. A variety of characteristics were significantly associated with a higher probability of receiving a penalty in HRRP (Table 2.3). Larger, teaching, non-profit, and northeast located hospitals all had increased odds of receiving a penalty when compared to their counterparts in bivariate specifications. In multivariate analyses, hospitals with 200-500 beds were still more likely to receive a penalty compared to small (<200 bed) hospitals. The relationship between teaching status and ownership became less significant in the multivariate specification, while hospitals in specific regions (West, Midwest, or South) were significantly less likely to receive a penalty.

Table 2.3 - Hospital Characteristics by Penalty in FY 2016

		Overall Sample	Penalty in FY 2016		Bivariate OR of Receiving Penalty in FY 2016 (95% CI)	Multivariate-Adjusted OR of Receiving Penalty in FY 2015 (95% CI)
			No	Yes		
Bed Size	<200	60%	20%	80%	Ref	Ref
	200-349	21%	13%	88%	1.79 (1.39 - 2.31)	1.62 (1.25-2.11)
	350-500	10%	12%	88%	1.88 (1.31-2.71)	1.69 (1.15-2.47)
	500+	9%	13%	87%	1.72 (1.18-2.49)	1.04 (0.63-1.72)
Teaching Status	None	64%	19%	81%	Ref	Ref
	Minor	28%	15%	85%	1.37 (1.10-1.70)	1.20 (0.94-1.52)
	Major	8%	9%	91%	2.38 (1.51-3.74)	1.99 (1.08-3.67)
Ownership	For Profit	23%	19%	81%	Ref	Ref
	Non Profit	62%	16%	84%	1.15 (0.92-1.44)	1.04 (0.82-1.31)
	Government	15%	18%	82%	1.00 (0.74-1.35)	0.96 (0.71-1.30)
Region	Northeast	16%	9%	91%	Ref	Ref
	West	19%	25%	75%	0.31 (0.21-0.44)	0.34 (0.24-0.50)
	Midwest	24%	20%	80%	0.40 (0.28-0.57)	0.43 (0.30-0.62)
	South	42%	15%	85%	0.56 (0.40-0.79)	0.67 (0.47-0.96)
	% Medicare Patients	40%	36%	40%	-	-

Table 2.4 provides information on the distribution of the excess readmission ratios in FY 2016 for the three conditions we use in our analysis. As expected, these distributions are centered around 1.00.

Table 2.4 - FY 2016 Excess Readmission Ratio Details

Excess Readmission Ratios	Hospitals	Mean	Standard Deviation	Minimum	Maximum
Acute Myocardial Infarction	2144	1	0.074	0.74	1.25
Heart Failure	2932	1	0.08	0.72	1.46
Pneumonia	2977	1	0.068	0.78	1.28

We have organized the remainder of the results by the independent variable of interest, and then show results when including all pertinent variables in the regression analyses. For each independent variable, we show results from a bivariate analysis in Model 1, and then from a regression that includes the hospital characteristics in Model 2 (bed-size, teaching status, ownership, and region). We show the main results when specifying the outcome variables of interest as the change in ERR from FY 2014 to FY 2016 whenever the independent variable of interest is observed in FY 2014. This discrepancy leads us to separate the results by variable of interest to maintain readability. Tables with the results when the outcome variable is the change in ERR from FY 2015 to FY 2016 are available in the Appendix. As a reminder, a negative outcome variable means an improvement over time in performance. Therefore, a negative coefficient on a variable implies that the variable was positively correlated with an improvement in performance.

Proportion of Medicare Patients

For every one percentage point increase in the Medicare patient proportion, on average, hospitals reduced their Heart Failure ERR by 0.0003, as seen in Table 2.5. But this result became insignificant when controlling for hospital characteristics. For the other two conditions, we find

no discernible correlation between proportion of Medicare patients and hospital improvements over time, both with and without hospital controls.

Table 2.5 - CMS Medicare Proportion and ERR Changes (FY16 - FY 14)

	AMI		**Heart Failure**		**Pneumonia**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
Medicare Days as a Percent of Total Inpatient Days (%)	-0.00007 (0.00012)	0.00008 (0.00015)	-0.00029** (0.00010)	-0.00020 (0.00011)	-0.00015 (0.00010)	-0.00014 (0.00010)
N	2068	2068	2916	2916	2957	2957

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses below coefficients.

While the directions of the coefficients were in the expected direction, the results are insignificant. Results from the FY 2016 – FY 2015 specification are seen in Table 2A.1, and parallel the insignificant results from above. From these findings, the evidence does not support our hypothesis as it seems that the proportion of inpatient days accounted for by Medicare patients is not associated with hospital improvement over time.

In specifications using the quartiles of the distribution, we find that the hospitals in the highest quartile of the Medicare proportion distribution improved their heart failure and pneumonia ERRs more than hospitals in the first quartile, but this result was insignificant when hospital characteristics were added (Appendix Table 2A.2).

Program Condition DRG Weight

We find that the relationship between the proportion of total DRG weights accounted for by the conditions measured in the program and improvements in performance over time is not

significant when controlling for hospital characteristics in the three original conditions (Table 2.6).

Table 2.6 - Program Condition DRG Proportion and ERR Changes (FY16 - FY 15)

	AMI		**Heart Failure**		**Pneumonia**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
Proportion of Total DRG Weights Accounted by Program Conditions (%)	0.00004 (0.00022)	0.00036 (0.00028)	-0.00016 (0.00016)	-0.00002 (0.00019)	0.00032* (0.00016)	0.00023 (0.00018)
N	2102	2102	2928	2928	2972	2972

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors reported in parentheses.

Only when we do not control for hospital characteristics, do we get a small effect in the Pneumonia ERR. In this situation, every percentage point increase in the proportion of total DRG weights accounted for by the 5 program conditions was correlated with hospitals increasing their ERR by 0.0003 points. In our alternative specifications, we find that even when specifying the DRG proportion as a categorical variable, the changes in excess readmission ratios did not significantly differ across these categories for most conditions. We only see a small effect when comparing the 26-50th percentile to the 0-25th percentile in Pneumonia (Appendix Table 2A.3).

Since CMS did not provide DRG information until FY 2015, we are unable to see how the results differ for the changes between FY 2014 and FY 2016. Evaluating this relationship among the two conditions that were added in FY 2015, we see that even after controlling for hospital characteristics, higher DRG proportions led to significant reductions in the excess readmission ratio for COPD (see Table 2.7). A ten percentage point increase in the DRG proportion would be associated with a 0.0049 decrease in the excess readmission ratio for COPD.

Table 2.7 - Program Condition DRG Proportion and New Condition ERR Changes (FY16 - FY 15)

	Hip/Knee Replacement		COPD	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*
Proportion of Total DRG Weights Accounted by Program Conditions (%)	-0.00044 (-0.00023)	-0.00031 (-0.00026)	-0.00047*** (-0.00014)	-0.00049** (-0.00017)
N	2102	2102	2898	2898

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors reported in parentheses.

Marginal benefits of improving performance in each condition

Even when controlling for hospital characteristics, we find that hospitals that have a higher marginal effect on their HRRP Adjustment Factor improve their excess readmission ratio significantly more than hospitals with smaller calculated marginal benefits (Table 2.8; model 1 not shown).

Table 2.8 - Calculated AF Marginal Benefit on ERR Changes (FY16 - FY 15)

	AMI	***HF***	***Pneumonia***	***Hip/Knee***	***COPD***
Variable	*Model 2*	*Model 2*	*Model 2*	*Model 2*	*Model 2*
Marginal Effect on Adjustment Factor in FY 15	-17.53*** (5.06)	-14.55*** (1.58)	-13.22*** (1.58)	-10.55*** (2.08)	-8.69*** (1.99)
N	879	1334	1348	1398	1723

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses below each coefficient. AF = Adjustment Factor

For every unit increase in the marginal change in the Adjustment Factor conditional on a decile improvement in performance, hospitals would significantly improve their excess

readmission ratios. To translate this into a more understandable scale for the marginal effects, a one standard deviation increase in the marginal effect of a decile improvement in performance was associated with a 0.0078 improvement in a hospital's AMI excess readmission ratio between FY15 and FY 16 (Table 2.8). The coefficients are smaller for the two most recent conditions. Since the effect of a marginal improvement has two downstream effects, changing the condition-specific excess readmission ratio as well as the HRRP Adjustment Factor, we also evaluate whether changes in ERR's over time are a function of the marginal benefit across all five conditions between FY 2015 and FY 2016.

Table 2.9 - Effect of Calculated ERR Marginal Benefit on Actual ERR Changes (FY16 - FY 15)

	AMI		**Heart Failure**		**Pneumonia**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
Marginal Effect of Improvement in Performance on ERR	-0.138*** (-0.031)	-0.139*** (-0.032)	-0.106*** (-0.020)	-0.103*** (-0.020)	-0.0769** (-0.027)	-0.078** (-0.027)
N	2094	2094	2903	2903	2946	2946

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses below each coefficient.

Again, we find that hospitals with higher marginal benefits of performance improvement on their condition-specific excess readmission ratios have significantly better changes than hospitals with less to gain from improvements in performance. A one standard deviation increase in the marginal effect of a decile improvement in performance was associated with a 0.0043 improvement in a hospital's heart failure excess readmission ratio (Table 2.9). The magnitude of improvement differed across the conditions, with higher effects seen in AMI than Pneumonia.

Table 2.10 -Effect of Calculated ERR Marginal Benefit on Actual ERR Changes (FY16 - FY 15)

	Hip/Knee		COPD	
Variable	Model 1	Model 2	Model 1	Model 2
Marginal Effect of Improvement in Performance on ERR	0.022 (0.037)	0.024 (0.037)	-0.065** (0.023)	-0.065** (0.023)
N	2420	2420	2873	2873

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses below coefficients.

For both AMI and heart failure, hospitals with the greatest benefit from a decile improvement in their performance made significantly greater improvements to their excess readmission ratios. This coefficient was smaller for both pneumonia and COPD, and for hip/knee arthroplasty, hospitals that had higher marginal benefits from an improvement in performance were not making significant improvements in their performance compared to hospitals with lower marginal benefits (Table 2.10). In the COPD condition, a one standard deviation increase in the marginal benefit of improvement was associated with a reduction in the ERR of -0.0021. The results seen for the Hip/Knee condition fall under the expectations of Hypothesis 5, as this condition was only added to the program in FY 2015, so hospitals may still be figuring out how best to improve their readmissions performance for patients with this condition. The differences in the Hip/Knee results between Table 2.8 and Table 2.10 can be attributed to the differences in the samples for each specification.

Our results from FY 2014 to FY 2016 analyses again show that hospitals with higher marginal benefits on the excess readmission ratio given a standardized marginal improvement

improved their excess readmission ratios significantly more than hospitals with smaller marginal benefits (Appendix Table 2A.4).

Previous Year Program Performance

To explore how the previous performance in the program correlated with changes in hospital performance over time, we used multiple specifications of prior performance. Hospitals receiving penalties in fiscal year 2014 of the program improved their performance significantly more compared to hospitals that avoided a penalty (Table 2.11). To provide a sense of the magnitude, more than 530 hospitals had a Pneumonia ERR that was between 1 and 1.036 in FY 2014, so if these hospitals had experienced this type of change, they would have avoided being penalized for the Pneumonia condition.

Table 2.11 - Effect of FY 14 Penalty on ERR Changes (FY16 - FY 14)

	AMI		**Heart Failure**		**Pneumonia**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
Received a Penalty in FY 14	-0.0282*** (0.0033)	-0.0290*** (0.0034)	-0.0332*** (0.0026)	-0.0338*** (0.0027)	-0.0353*** (0.0024)	-0.0355*** (0.0025)
N	2068	2068	2916	2916	2957	2957

Note: * for p<.05, ** for p<.01, and *** for p<.001

For hospitals that received a penalty in FY 2015, they were more likely to improve their performance on ERR's for all three conditions (Appendix Table 2A.5). For example, hospitals that received a penalty improved their pneumonia excess readmission ratios by 0.015 compared to those hospitals that avoided a penalty, even when controlling for hospital characteristics. The magnitude of the effect varied slightly across the conditions.

These results suggest that hospitals receiving a penalty may be more likely to improve their performance, though we cannot make any causal inferences. The increased magnitude of the coefficients in fiscal year 2014 is consistent with the notion that the timing of FY 2014 results allows hospitals to make changes to their performance for the FY 2016 program. Yet, since the general trend of the program has been improvement over time, it is difficult to tie this improvement specifically to penalty receipt alone. Turning to how the actual excess readmission ratio from a previous fiscal year may or may not be correlated to changes in hospital performance, we find that worse performance in previous years was significantly associated with improvements over time.

A 0.01 increase in the pneumonia ERR in FY 2014 was associated with a 0.00487 improvement over the two fiscal years (Table 2.12). This effect was seen across all three conditions, where higher previous excess readmission ratios were associated with significant improvements between FY 2014 and FY 2016. This relationship was also seen in the FY 2015/FY 2016 analyses (Appendix Table 2A.6). Every 0.01 increase in the excess readmission ratio for AMI was correlated with a 0.0018 improvement in ERR's from FY 2015 to FY 2016. Parallel to the trend discussed above, the coefficient magnitudes were all larger in the FY 2014-FY 2016 specifications.

Table 2.12 - Effect of Previous ERR on ERR Changes (FY16 - FY 14)

	AMI		**Heart Failure**		**Pneumonia**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
ERR in FY 2014	-0.460*** (-0.018)	-0.480*** (-0.018)	-0.391*** (0.015)	-0.408*** (-0.015)	-0.476*** (-0.014)	-0.487*** (-0.014)
N	2068	2068	2916	2916	2957	2957

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses below coefficients.

It is important to note that these effects are averaged out over the whole distribution of each of the excess readmission ratios. Since there is a non-linear component to the incentive structure, as ERR's below 1 help hospitals reduce or avoid penalties, we wanted to determine what the effect was of a jump in ERR's from below 1 to above 1. Hospitals with a heart failure ERR above 1 in FY 2014 improved their performance by 0.0534 more than hospitals that had an ERR less than or equal to 1 (see Table 2.13). The effects were similar for all conditions, even after controlling for hospital characteristics. Again, as seen before, the magnitude was larger when looking at hospital performance in FY 2014. Our results for FY 2015 are seen in Appendix Table 2A.7. From these results, a hospital with an ERR above 1 for AMI improved their performance by 0.021 more than a hospital with an ERR less than or equal to 1.

Table 2.13 - Effect of ERR Above 1 on ERR Changes (FY16 - FY 14)

	AMI		**Heart Failure**		**Pneumonia**	
Variable	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
ERR Was Above 1 in FY 14	-0.0577*** (0.0029)	-0.0594*** (0.0029)	-0.0534*** (0.0024)	-0.0546*** (0.0024)	-0.0596*** (0.0022)	-0.0602*** (0.0022)
N	2068	2068	2916	2916	2957	2957

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors are in parentheses below the coefficients.

Thus, penalty receipt seems to be correlated with a significant improvement in excess readmission ratios over time. This confirms our hypothesis that hospitals receiving a penalty would have significantly larger improvements in their performance compared to those that avoided a penalty. Further, having condition-specific ERR's above 1, which is an important determinant of penalty receipt, was also significantly correlated with improvements in performance over time.

We also ran regressions where the previous year ERR was specified in a piecewise form in the regression, allowing for a different slope when a previous ERR was above 1. We specified a break in the previous year ERR right at 1, so the coefficient on the ERR spline being above 1 shows the change in slopes from below 1 to above 1. Evaluating the results for FY 2014, we see that there was a significant change in the slopes, with the ratios above 1 leading to a significantly greater change in performance compared to hospitals below 1 (Table 2.14).

Table 2.14 – Additive Piecewise ERR Regression (FY16 - FY 14)

	AMI		**Heart Failure**		**Pneumonia**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
ERR Spline <=1	-0.383*** (-0.032)	-0.400*** (-0.032)	-0.339*** (-0.027)	-0.353*** (-0.027)	-0.424*** (-0.026)	-0.435*** (-0.026)
ERR Spline > 1	-0.151** (-0.057)	-0.161** (-0.057)	-0.098* (-0.048)	-0.106* (-0.049)	-0.095* (-0.044)	-0.094* (-0.044)
N	2068	2068	2916	2916	2957	2957

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors are in parentheses below coefficients. The spline above 1 shows effects that are additive to the 1st coefficient.

Hospitals with an ERR above 1 had a 0.106 greater effect on changes in performance than those with ERR's less than or equal to 1 for heart failure. The results from FY 2015 suggest that there was not a significant change in the slope of the effect of an ERR above 1 when compared to below 1 for all three conditions (Appendix Table 2A.8). This difference in significance of the jumps in the slopes from below to above 1 between FY 2014 to FY 2015 aligns with the fact that hospitals have time to review their performance in fiscal year 2014 and make changes to their performance that would impact their scoring in fiscal year 2016.

Finally, we wanted to disentangle whether the penalty size was correlated with hospital performance compared to general penalty avoidance. To do so, we limit our analysis to hospitals with ERR's above 1, and categorize them into groups of 25th percentiles in the ERR distribution

above 1. If the size of the penalty drives hospital behavior to a greater degree than avoiding penalties, then hospitals in the 4th group (75th-100th percentile) should have a greater improvement than hospitals in the 0-25th percentile.

Table 2.15 - ERR Performance Above 1 (FY16 - FY 14)

	AMI		**Heart Failure**		**Pneumonia**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
26-50th Percentile	-0.0235*** (0.0055)	-0.0229*** (0.0055)	-0.0093* (0.0045)	-0.0102* (0.0045)	-0.0102* (0.0042)	-0.0109** (0.0042)
51-75th percentile	-0.0302*** (0.0055)	-0.0307*** (0.0054)	-0.0172*** (0.0048)	-0.0195*** (0.0047)	-0.0280*** (0.0043)	-0.0299*** (0.0043)
76-100th percentile	-0.0640*** (0.0058)	-0.0666*** (0.0058)	-0.0531*** (0.0051)	-0.0565*** (0.0051)	-0.0656*** (0.0045)	-0.0677*** (0.0046)
N	1029	1029	1439	1439	1411	1411

Note: * for p<.05, ** for p<.01, and *** for p<.001. Coefficient with standard error below.

As can be seen above, hospitals that did the worst (75th-100th percentile of the ERR distribution above 1) improved their performance significantly more when compared to hospitals in the 0-25th percentile. This result is in terms of raw improvements in the ERR, so while the hospitals right above 1 may be able to avoid a penalty in a future year, those that performed the worst were able to improve their ERR's significantly more. This goes counter to the expectation of other researchers that the worst performers will find it less likely to work on improving their chances of avoiding a penalty (Rosenthal et al., 2004). For hospitals that were in the 76th-100th percentile of the distribution of ERR's above 1 for AMI in FY 2014, they improved their ERR by 0.0666 more than hospitals in the 0-25th percentile. This magnitude and significance was found across all three conditions evaluated. Our results were confirmed in our FY 2015 specifications

(Appendix Table 2A.9) as the worst-performing hospitals were able to improve significantly more than hospitals right above 1 across all three conditions.

Including all independent variables of interest into one specification allows us to get a glimpse into how hospitals may react while taking into consideration all these characteristics. If all these variables are important, including them all in one specification provides unbiased estimates of their effects on changes in hospital performance over time. We find that the variable that is always significant is the marginal benefit of improvements (Table 2.16).

Table 2.16 - Correlation of Changes in ERR with Sources of Incentive Heterogeneity (FY16 - FY 15)

Variable	**AMI**	**Heart Failure**	**Pneumonia**
Medicare Days as a Percent of Total Inpatient Days (%)	0.00003 (0.00009)	-0.00003 (0.00007)	-0.00016 (0.00007)
Penalty in FY 2015	-0.0032 (0.0031)	-0.0080*** (0.0023)	-0.014*** (0.002)
Proportion of Total DRG Weights Accounted by Program Conditions (%)	-0.00022 (0.00029)	-0.00013 (0.00020)	0.00016 (0.00019)
Marginal Benefit on FY 2015 ERR	-0.13*** (0.03)	-0.089*** (0.021)	-0.057*** (0.027)
N	2,094	2,903	2,946

Note: * for p<.05, ** for p<.01, and *** for p<.001. Controlling for hospital bed size, ownership, teaching status, and census region.

For both heart failure and pneumonia, the receipt of a penalty also is significant in predicting changes in hospital excess readmission ratios. This result is expected when we include the penalty receipt, as there is a high degree of correlation between previous performance and the marginal benefit of a decile improvement. Thus, we see that the marginal benefit of

improvements in performance, is the main determinant of changes in excess readmission ratios over time. As not all variables of interest were observed in FY 2014, we do not include them in the main results here. We run as full a specification as possible when studying the changes in ERR's from FY 2014 to FY 2016 and find that the previous excess readmission ratio was again the only component that was significantly correlated with improvement (see Appendix Table 2A.10).

Condition Specific Responses

In analyzing whether program timing prevents hospitals from making improvements in readmission ratios in subsequent years, we compare the correlation of our independent variables of interest with changes in the excess readmission ratios between FY 2016 and FY 2015 for Hip/Knee Replacement as well as COPD. We find that even for conditions that were only added to the program in FY 2015, receiving a penalty in FY 2015 is significantly associated with improvements in the ratio (Table 2.17). This correlation, along with the knowledge that hospitals did not have time to see their fiscal year 2015 results and make improvements in their performance levels that would be captured in the fiscal year 2016 program suggest that hospitals participating in HRRP may be applying their methods to reduce readmissions on conditions beyond the original three incentivized conditions.

The newly added conditions are the only ones where we see significant relationships between changes in performance and program DRG proportion. This could be driven by the fact that as more conditions get added to the program, hospitals do have to make conscious decisions regarding whether or not to engage in quality improvement efforts for their readmissions performance. Yet, more work is needed to identify causal mechanisms.

Table 2.17 - Correlation of Changes in New Condition ERR's with Sources of Heterogeneity (FY16 - FY 15)

Variable	**Hip/Knee Replacement**	**COPD**
Medicare Days as a Percent of Total Inpatient Days (%)	0.00009 (0.00016)	0.00007 (0.00006)
Penalty in FY 2015	-0.023*** (0.005)	-0.005*** (0.002)
Proportion of Total DRG Weights Accounted by Program Conditions (%)	-0.0007** (0.0003)	-0.0006*** (0.0002)
Marginal Benefit on FY 2015 ERR	-0.04*** (0.04)	-0.060** (0.023)
N	2420	2873

Note: * for p<.05, ** for p<.01, and *** for p<.001. Controlling for hospital bed size, teaching status, ownership, and census region. Standard error in parentheses below coefficients

Robustness Check

The issue of endogeneity is unavoidable in an analysis of hospital behavior in an incentive program. This issue is brought about even more since the program was introduced at once to all participating hospitals, reducing our ability to make causal inferences in a straightforward manner. One possible route for biased results is whether payer mix for hospitals is a driver of quality improvement efforts. Hypothetically, a poor-performing hospital could react to reductions in their Medicare reimbursements by refusing to accept Medicare, or changing their coding patterns to make up for any lost revenue caused by a penalty. The notion that hospitals alter their coding behavior to improve reimbursement rates has been studied for decades, with documentation of coding error rates as high as 20.8% in some studies (Farmer et al., 2013; Hsia

et al., 1988; Silverman and Skinner, 2004). Since we are evaluating hospitals in the fourth fiscal year of the program, if they have made changes to their accepted patient population, or their coding habits, then both these coefficients could be biased. As pointed out by Sloan et al., these responses should depend on the ownership status of a hospital, as a profit-seeking firm will engage in behaviors to maximize their reimbursements (Sloan et al., 2001).

Previous research into hospital behavior after reductions in Medicaid reimbursements found reductions in services per admission for Medicaid patients, with the reduction more significant for Medicaid patients than patients covered by other sources (Dranove and White, 1998). Given that thus far in the program, there have not been many hospitals that have been able to avoid penalties in subsequent years, hospitals with a high proportion of Medicare patients may be shifting their resources elsewhere instead of responding to the incentives (Dranove and White, 1998). As Feder et al. outline in their paper, if quality is a private good, then reductions in Medicaid reimbursements would decrease quality for Medicaid patients, and increase quality for non-Medicaid patients. On the other hand, if quality is a public good, then reductions in Medicaid reimbursements would lead to reduced quality for Medicaid patients and potential decreases in quality for non-Medicaid patients (Feder et al., 1987; Dranove and White, 1998). While these conclusions are based on situations where the financial cuts were not driven by quality performance, the theory does suggest that there may be an underlying relationship between payer mix, reimbursement changes, and the propensity to improve quality for patients. Furthermore, As Norton and Staiger find, hospital ownership can be endogenous to the volume of uninsured patients, so the relationship between ownership and payer mix may also drive quality improvement efforts (Norton and Staiger, 1994). Other work has shown decreases in Medicare cases and length of stay after the introduction of the prospective payment system,

which significantly changed the reimbursement levels for hospitals (Feder et al., 1987). But it is not necessarily an absolute truth that cuts to Medicare reimbursements leads to decreased quality for Medicare patients, as Seshamani et al. found no differences in mortality for hospitals that suffered large Medicare reimbursement cuts as a result of the Balanced Budget Act of 1997 (Seshamani et al., 2006). While many of the analyses done thus far have found negative impacts of financial shocks on quality of care, the difference in the cuts to reimbursements hospitals face in the context of the HRRP is that the cuts are not exogenously applied.

We only observe the DRG weights for two years, but we analyzed them to determine if there is a relative consistency of DRG proportions across fiscal years 2015 and 2016. Since the program relies on a 3-year moving period of data, there are only a couple of months of claims that differentiate these two fiscal years. Additionally, we compare the proportion of inpatient days accounted for by Medicare across multiple years to see if there is a trend in either direction for hospitals. Between FY 2011 and FY 2016 there was a decrease in the proportion of inpatient days accounted for by Medicare (Table 2.18). Hospitals that avoided a penalty in FY 2013 or FY 2015 had a significantly larger decrease in their share of inpatient days accounted for by Medicare than hospitals that received penalties in these two fiscal years. In terms of the condition DRG weights, we find that hospitals that avoided a penalty in FY 2015 actually increased their condition DRG proportion between the two years significantly more than hospitals that received a penalty in FY 2015 (Table 2.19).

Table 2.18 - Difference in Medicare Proportion of Inpatient Days by HRRP Penalty Status

		FY 2015 - FY 2011	p-value
FY 2013	Penalty	-0.053	0.0027
	No Penalty	-0.064	
FY 2014	Penalty	-0.059	0.67
	No Penalty	-0.061	
FY 2015	Penalty	-0.052	0.0074
	No Penalty	-0.062	

Table 2.19 - Difference in HRRP Condition DRG Percent by FY 15 Penalty Status

		HRRP DRG Weights FY15	**HRRP DRG Weights FY16**	**p-value difference**
FY 2015	No Penalty	15.6%	17.6%	<0.0001
	Penalty	15.5%	15.8%	

Another possible explanation driving the magnitude of results we observe is that the changes in excess readmission ratios are driven by hospital regression to the mean over time in terms of readmissions performance. To identify what may be the magnitude of the regression to the mean change in performance over time, we calculated the average change in excess readmission ratios between FY 2014 and FY 2016 for hospitals that were persistently high performers in a condition. For those hospitals that are always in the top decile of excess readmission ratio performance in a specific condition, their overall performance in the condition does not vary, but their raw excess readmission ratios may move closer to the mean over time. If

these hospitals actually decrease their excess readmission ratio over time, then that would suggest that the threshold for reducing preventable readmissions may not have been reached yet. On the other hand, if these persistently high-performing hospitals experience an increase in their excess readmission ratios, then the magnitude of that increase could represent the movement towards the true mean performance in readmissions.

We see that for the hospitals that were in the top decile of ERR's in both FY 2014 and FY 2016 for each of the three original conditions, they all experienced an increase in their ERR over time (see Table 2.20). For the 123 hospitals that were in the top decile of excess readmission ratios for pneumonia in both FY 2014 and FY 2016, the average change in the ERR over this time was an increase by 0.015. In the other two conditions, this increase was slightly smaller, but still showed that over time these persistently high performers increased their excess readmission ratio. If these values capture some of the regression to the mean of hospital readmissions performance, they still are smaller than the magnitudes of many of the coefficients. Therefore, regression to the mean does not alone explain the trends in hospital improvements over time in the Hospital Readmissions Reduction Program.

Table 2.20 - Regression to the Mean Calculations for Top Performing Hospitals

Condition	*Hospitals*	*Average ERR Change (FY 16-FY14)*
AMI	88	0.009
HF	146	0.006
PN	123	0.015

Limitations

The analyses are limited by program structure as we are unable to make causal inferences without a true control sample of hospitals. As the program was rolled out to many acute care hospitals at the same time, there is no ideal control group for our analyses. Further, without DRG data for the first year of our data analyses, we cannot analyze the marginal benefit of improvements in performance from FY 2014 results on the HRRP Adjustment Factor. While our calculations allow us to determine the effect of a standardized improvement in FY 2014 on the ERR, the main input in determining hospital performance in the program, linking these changes to overall program performance would have been more relevant for hospitals. Our analyses only look at the output of hospital performance (readmissions reduction), and we do not capture the quality improvement resources or other inputs hospitals have invested in to work towards improving their quality. We also cannot precisely replicate the HRRP Adjustment Factor for all hospitals in our sample, which seems to be related to the data released by CMS. This led to differences in results in our marginal effects analyses for the Hip/Knee arthroplasty condition, which merits further evaluation. The heterogeneity in the costs of quality improvement investments could be an important driver in the effort hospitals put forth to improve their performance. Qualitative analyses using structured interviews could help identify other hospital characteristics that are associated with response to incentives, as previous work has identified some factors that seem to motivate performance improvement (Reiter et al., 2006). Finally, due to data limitations, we use the percent of inpatient days Medicare accounts for as our proxy of Medicare reimbursement proportion at the hospital level. With more detailed hospital financial information, we would be able to discern the true proportion of dependency a hospital has on Medicare-based reimbursements.

Every year, there are some programmatic changes that take place for the hospital quality incentive programs administered by CMS. Beginning in FY 2017, readmissions after coronary artery bypass graft surgery will also be included in the program calculations. With changes in included diagnoses and their preventable readmission calculations, how hospitals perform and react to their performance may change in the future. With complete claims-level data, we would be able to replicate the exact hospital-specific effect for each hospital in each fiscal year, and be able to confirm our marginal benefit calculations. Finally, the timing of the program makes a comparison of FY 2014 and FY 2016 performance levels more accurate than subsequent years. Unfortunately, the DRG proportion was not available for FY 2014.

Discussion

We find varying results when evaluating whether or not drivers of incentive heterogeneity for hospitals participating in the Hospital Readmissions Reduction Program are correlated with changes in performance over time. While hospitals did not improve to a greater degree based on their proportion of Medicare patients, we do find that hospitals are responding to the main incentive in the program. Hospitals that either received a penalty or performed poorly compared to the other hospitals in the program had significantly greater improvements over time in their readmissions performance. In this regard, the CMS program is having its intended effect – penalties are incentivizing them to work on improving their quality as measured by reducing preventable readmissions. As CMS continues to add conditions to the program, the proportion of hospitals receiving penalties could rise, and the effectiveness of a penalty in driving improvements may change. Hospitals also are making changes in conditions where the marginal benefit of improvement is the highest, such that hospitals with more to gain by improving their

performance in a specific condition have significantly greater improvements than hospitals that have smaller marginal benefits.

While we do find evidence of regression to the mean as hospitals that persistently performed very well in condition-specific readmission rates actually experienced increases in their excess readmission ratio over time, the magnitude of this increase was smaller than the improvements in performance we find for hospitals that received penalties or that had extremely poor excess readmission ratios. Another contribution of this analysis is the determination of the marginal benefits of a standardized level of improvement in performance on program performance for hospitals. Our measure of standardized improvements translated into 2-4 less readmissions on average within each condition. Since hospitals face a variety of resource and time constraints, maximizing quality investments so that they are able to reap the greatest benefit in future years would be a reasonable goal. Our results confirm our hypothesis, as we find that hospitals do seem to be improving their performance when their marginal benefits are higher in most conditions. As CMS continues to add eligible conditions to the scoring methodology of the Hospital Readmissions Reduction Program, hospitals able to identify the conditions with the greatest marginal benefit could help reduce or avoid penalties altogether. Our analysis finds that not all hospitals with the most to gain from improvements in readmissions for patients of hip/knee arthroplasty are not taking advantage of this opportunity yet.

Beyond differences in magnitudes of coefficients across diseases, we find that for the recently added conditions, program-condition DRG proportion is significantly related to changes over time. For these two conditions, this result matches our hypotheses – an increased proportion of overall services accounted for by the conditions in the program would be correlated with better improvements over time. As more conditions continue to be added to the program, this

value could increase in importance. Having to improve readmissions across six conditions may be a different endeavor compared to just three, so the service volumes of each of these conditions could change how hospitals react to the multiple-condition incentive program. Future research with more years of data may be able to identify whether the relationship between the program condition DRG proportion is increasing in importance.

Although many studies have documented the relative success of the Hospital Readmissions Reduction Program in terms of improving targeted as well as all-cause readmission rates in participating hospitals over the last couple of years, to our knowledge this is the first analysis that attempts to identify the incentive levers inherent in the program that could be pushing hospitals to improve. In our nonexperimental study design, we identify associations between some of the sources of incentive heterogeneity in the program and hospital performance over time. Even though we cannot make causal inferences regarding these different sources of heterogeneity, we believe that the information is valuable nonetheless. Hospitals do not incorporate reliance on insurance reimbursements in their decisions of whether or not to focus on readmissions improvement. On the other hand, the proportion of a hospital's services directly measured in the program is related to performance improvements in the two conditions added in FY 2015. Furthermore, penalty receipt and poor performance compared to others are highly correlated with improvements over time across all five conditions. With our novel methodology to identify the marginal benefit of improvements on program outcomes, we find that for most conditions, hospitals do seem to be improving where there is the greatest benefit.

References

Centers for Medicare & Medicaid Services, Federal Register 2012b. Medicare program; hospital inpatient prospective payment systems for acute care hospitals and the long-term care hospital prospective payment system and fiscal year 2013 rates; hospitals' resident caps for graduate medical education payment purposes; quality reporting requirements for specific providers and for ambulatory surgical centers, Final Rule, 77, 53257-53750.

Abelson R. 2013. Hospitals question Medicare rules on readmissoins. *New York Times.*

Anderson GF, Steinberg EP,. 1984. Hospital readmissions in the Medicare population. New England Journal of Medicine 311; 1349-1353.

Ashton CM, Kuykendall DH, Johnson ML, Wray NP, Wu L,. 1995. The association between the quality of inpatient care and early readmission. Ann Intern Med 122; 415-421.

Bartel A, Chan, CW, Kim, SH,. 2014. Should hospitals keep their patients longer? The role of inpatient and outpatient care in reducing readmissions. . National Bureau of Economic Research Working Paper Series 20499.

Berenson RA, Paulus RA, Kalman NS,. 2012. Medicare's readmissions-reduction program — a positive alternative. New England Journal of Medicine 366; 1364-1366.

Berwick DM, Hackbarth AD,. 2012. Eliminating waste in us health care. JAMA 307; 1513-1516.

Boccuti C, & Casillas, G. 2015. Aiming for fewer hospital u-turns: the Medicare hospital readmission reduction program. *Kaiser Family Foundation*.

Boozary AS, Manchin J, Iii, Wicker RF,. 2015. The medicare hospital readmissions reduction program: Time for reform. JAMA 314; 347-348.

Dorsey K, et al. **2015**. Condition-Specific Measures Updates and Specifications Report: Hospital-Level 30-Day Risk-Standardized Readmission Measures. Centers for Medicare & Medicaid Services; 2015.

Dranove D, White WD,. 1998. Medicaid-dependent hospitals and their patients: how have they fared? Health Services Research 33; 163-185.

Farmer SA, Black B, Bonow RO,. 2013. Tension between quality measurement, public quality reporting, and pay for performance. JAMA 309; 349-350.

Feder J, Hadley J, Zuckerman S,. 1987. How did Medicare's prospective payment system affect hospitals? New England Journal of Medicine 317; 867-873.

Friedman B, Basu J,. 2004. The rate and cost of hospital readmissions for preventable conditions. Medical Care Research and Review 61; 225-240.

Gerhardt G, Yemane A, Hickman P, Oelschlaeger A, Rollins E, Brennan N,. 2013. Medicare readmission rates showed meaningful decline in 2012. Medicare Medicaid Res Rev 3.

Goldfield NIMD, McCullough ECMS, Hughes JSMD, Tang AM, Eastman BMS, Rawlins LK, Averill RFMS,. 2008. Identifying potentially preventable readmissions. Health Care Financing Review 30; 75-91.

Hines A, Barrett, ML, Jiang HJ, and Steiner, CA,. 2014. Conditions with the largest number of adult hospital readmissions by payer, 2011. HCUP Statistical Brief #172. In: Quality AfHRa (Ed)^(Eds). Rockville, MD; 2014.

Hsia DC, Krushat WM, Fagan AB, Tebbutt JA, Kusserow RP,. 1988. Accuracy of diagnostic coding for Medicare patients under the prospective-payment system. New England Journal of Medicine 318; 352-355.

Joynt KE, Jha AK,. 2013. Characteristics of hospitals receiving penalties under the hospital readmissions reduction program. JAMA 309; 342-343.

Keefe A. 2016. Institute of Medicine Report Examines Socioeconomic Risk Adjustment. (Ed)^(Eds), vol. 2016. California Hospital Association.

McIlvennan CK, Eapen ZJ, Allen LA,. 2015. Hospital readmissions reduction program. Circulation 131; 1796-1803.

Menachemi N, Matthews MC, Ford EW, Brooks RG,. 2007. The influence of payer mix on electronic health record adoption by physicians. Health Care Management Review 32; 111-118.

Ness D, & Kramer, W,. 2013. Reducing hospital readmissions: It's about improving patient care. . (Ed)^(Eds), Health Affairs Blog, vol. 2015. Health Affairs.

Norton EC, Staiger DO,. 1994. How hospital ownership affects access to care for the uninsured. The RAND Journal of Economics 25; 171-185.

Reiter KL, Nahra TA, Alexander JA, Wheeler JR,. 2006. Hospital responses to pay-for-performance incentives. Health Services Management Research 19; 123-134.

Rosenthal MB, Fernandopulle R, Song HR, Landon B,. 2004. Paying For Quality: Providers' Incentives For Quality Improvement. Health Affairs 23; 127-141.

Centers for Medicare & Medicaid Services, 2013. Acute Care Hospital Inpatient Prospective Payment System.

Seshamani M, Schwartz JS, Volpp KG,. 2006. The effect of cuts in Medicare reimbursement on hospital mortality. Health Services Research 41; 683-700.

Silverman E, Skinner J,. 2004. Medicare upcoding and hospital ownership. Journal of Health Economics 23; 369-389.

Sloan FA, Picone GA, Taylor Jr DH, Chou S-Y,. 2001. Hospital ownership and cost and quality of care: is there a dime's worth of difference? Journal of Health Economics 20; 1-21.

Zuckerman RB, Sheingold SH, Orav EJ, Ruhter J, Epstein AM,. 2016. Readmissions, observation, and the hospital readmissions reduction program. New England Journal of Medicine 374; 1543-1551.

Appendix

Table 2A 1 - Effect of Medicare Proportion on ERR Changes (FY16 - FY 15)

	AMI		**Heart Failure**		**Pneumonia**		**Hip/Knee Replacement**		**COPD**	
Variable	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
Medicare Days as a Percent of Total Inpatient Days (%)	0.00003 (0.00008)	0.00001 (0.00009)	-0.00006 (0.00006)	-0.00003 (0.00007)	-0.00008 (.00006)	-0.0001 (0.00007)	0.00001 (-0.00015)	0.00002 (-0.00016)	0.0000030 (-0.000053)	0.000027 (-0.000058)
N	2102	2102	2928	2928	2972	2972	2433	2433	2898	2898

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses below coefficients

Table 2A.2 – Effect of Medicare Proportion Distribution on ERR Changes (FY 16 – FY14)

Variables	**AMI**		**HF**		**PN**	
	Model 1	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
26-50th Percentile	-0.004 (0.004)	-0.001 (0.004)	0.0007 (0.004)	0.0019 (0.004)	-0.0016 (0.003)	-0.0012 (0.003)
51-75th Percentile	-0.003 (0.004)	0.00001 (0.005)	0.0019 (0.004)	0.0039 (0.004)	-0.0011 (0.003)	-0.0007 (0.004)
76-100th Percentile	-0.002 (0.005)	0.002 (0.005)	-0.0085* (0.004)	-0.0053 (0.004)	-0.0063 (0.003)	-0.0059 (0.004)

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses below coefficients. Reference group is the 0-25th percentile of the Medicare proportion distribution

Table 2A.3 - Effect of HRRP DRG Distribution on Changes in ERR (FY 16 - FY 15)

Variables	**AMI**		**Heart Failure**		**Pneumonia**	
	Model 1	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
26-50th Percentile	0.0035	0.0047	0.0033	0.0041	0.0059*	0.0058*
	(0.0027)	(0.0029)	(0.0024)	(0.0025)	(0.0023)	(0.0024)
51-75th Percentile	0.0009	0.0031	-0.0000	0.0017	0.0033	0.0031
	(0.0027)	(0.0031)	(0.0024)	(0.0026)	(0.0022)	(0.0024)
76-100th Percentile	-0.0006	0.0023	-0.0018	0.0004	0.0058*	0.0050
	(0.0031)	(0.0036)	(0.0024)	(0.0028)	(0.0023)	(0.0026)

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses below coefficients

Table 2A.4 - Effect of Calculated Marginal Benefit on ERR Changes (FY16 - FY 14)

	AMI		**Heart Failure**		**Pneumonia**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
Marginal Effect of Improvement in Performance on ERR	-0.248*** (0.038)	-0.251*** (0.039)	-0.157** (0.049)	-0.147** (0.050)	-0.172*** (0.039)	-0.169*** (0.040)
N	2067	2067	2913	2913	2954	2954

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses below coefficients.

Table 2A.5 - Effect of FY 15 Penalty on ERR Changes (FY16 - FY 15)

	AMI		**Heart Failure**		**Pneumonia**		**Hip/Knee**		**COPD**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
Received a Penalty in FY 15	-0.0064* (0.0029)	-0.0064* (0.0030)	-0.0105*** (0.0021)	-0.0105*** (0.0022)	-0.0153*** (0.0019)	-0.0150*** (0.0020)	-0.0161*** (0.0041)	-0.0184*** (0.0042)	-0.0048*** (0.0019)	-0.0052*** (0.0019)
N	2102	2102	2928	2928	2972	2972	2433	2433	2898	2898

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses.

Table 2A.6 - Effect of Previous ERR on ERR Changes (FY16 - FY 15)

	AMI		**Heart Failure**		**Pneumonia**		**Hip/Knee**		**COPD**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
ERR in FY 2015	-0.178*** (0.014)	-0.190*** (0.014)	-0.170*** (0.011)	-0.179*** (0.011)	-0.262*** (0.010)	-0.267*** (0.011)	-0.199*** (0.0133)	-0.204*** (0.013)	-0.153*** (0.011)	-0.161*** (0.011)
N	2102	2102	2928	2928	2972	2972	2433	2433	2898	2898

Note: * for p<.05, ** for p<.01, and *** for p<.001. Standard errors in parentheses.

Table 2A.7 - Effect of ERR Above 1 on ERR Changes (FY16 - FY 15)

	AMI		**Heart Failure**		**Pneumonia**	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
ERR Above 1 in FY 15	-0.0210*** (0.0020)	-0.0220*** (0.0020)	-0.0246*** (0.0016)	-0.0254*** (0.0017)	-0.0316*** (0.0015)	-0.0318*** (0.0016)
N	2102	2102	2928	2928	2972	2972

Note: * for p<.05, ** for p<.01, and *** for p<.001

Table 2A.8 - Piecewise ERR Regression (FY16 - FY 15)

	AMI		Heart Failure		Pneumonia	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
ERR Spline <=1	-0.179*** (0.025)	-0.190*** (0.025)	-0.172*** (0.019)	-0.179*** (0.020)	-0.234*** (0.020)	-0.240*** (0.020)
ERR Spline > 1	0.0005 (0.0436)	-0.004 (0.044)	0.004 (0.034)	-0.0006 (0.0022)	-0.051 (0.034)	-0.049 (0.034)
N	2102	2102	2928	2928	2972	2972

Note: * for p<.05, ** for p<.01, and *** for p<.001

Table 2A.9 - ERR Performance Above 1 (FY16 - FY 15)

	AMI		Heart Failure		Pneumonia	
Variable	*Model 1*	*Model 2*	*Model 1*	*Model 2*	*Model 1*	*Model 2*
26-50th Percentile	0.0025 (0.0039)	0.0019 (0.0039)	-0.0107** (0.0033)	-0.0110*** (0.0033)	-0.0068* (0.0030)	-0.0068* (.0030)
51-75th percentile	-0.0001 (0.0041)	0.0005 (0.0041)	-0.0112** (0.0034)	-0.0124*** (0.0034)	-0.0164*** (0.0030)	-0.0167*** (0.0031)
76-100th percentile	-0.017*** (0.0041)	-0.018*** (0.0041)	-0.0190*** (0.0035)	-0.0211*** (0.0036)	-0.0330*** (0.0032)	-0.0337*** (0.0033)
N	1035	1035	1427	1427	1394	1394

Note: * for p<.05, ** for p<.01, and *** for p<.001. Coefficient with standard error below.

Table 2A.10 - Correlation of Changes in ERR with Sources of Incentive Heterogeneity (FY16 - FY 14)

Variable	**AMI**	**Heart Failure**	**Pneumonia**
Medicare Days as a Percent of Total Inpatient Days (%)	0.000068 (0.00013)	-0.000140 (0.000097)	$-5.11*10^{-6}$ (0.000083)
ERR in FY 2015	-0.499*** (0.020)	-0.412*** (0.016)	-0.491*** (0.014)
Marginal Benefit on ERR	0.080* (0.033)	0.040 (0.042)	0.042 (0.028)
N	2,067	2,913	2,954

Note: * for p<.05, ** for p<.01, and *** for p<.001. Controlling for hospital bed size, ownership, teaching status, and census region. Standard errors in parentheses below coefficients.

Chapter 3
Adding Depression to a Microsimulation Model of Diabetes Progression

Introduction

As the prevalence of diabetes continues to grow in the United States and abroad, identifying effective treatment methods remains a paramount health care need. In 2012, there were an estimated 29.1 million individuals with diabetes in the United States [1]. This was an increase from 25.8 million in 2010. Approximately 1.4 million individuals are diagnosed with diabetes every year, though 8.1 of the 29.1 million cases in 2012 were undiagnosed [1]. While recent evidence suggests that we have been making some progress in slowing the increase of diabetes and improving diabetes management and diagnosis, there remain plenty of areas for improvement [2, 3]. Concurrent with the rising numbers of individuals with diabetes in the US is the cost of care for these patients. It is estimated that the US spent $176 billion on health care related to diabetes in 2012, with indirect costs adding another $69 billion [4]. As health care costs continue to be a source of concern, the implementation of interventions that help reduce costs and improve health are critically needed. A contributing factor to the high costs and management complexity of diabetes is the characteristic development of many comorbidities and complications over time. In a recent position statement of the American Diabetes Association and the European Association for the Study of Diabetes detailing a patient-centered approach to managing type 2 diabetes, the comorbidities listed included coronary artery disease, heart failure, chronic kidney disease, liver dysfunction, and hypoglycemia [5]. Each of these in isolation or in

combination with others necessitates greater health care management by both the physician and the patient.

While the aforementioned conditions co-occur with diabetes at an increased rate and merit further study, they are not the only conditions that can develop with diabetes. One commonly underdiagnosed and undertreated condition in this patient population is depression. Depression is one of the most prevalent mental health illnesses in the population, and when it remains undiagnosed or poorly treated, it can lead to death. The prevalence of depression among adults 18 years or older in 2014 was estimated to be 15.7 million, representing 6.7% of the population [6]. The direct and indirect medical costs associated with depression can be as high as $210.5 billion annually, when including costs from depression as well as comorbid conditions [7]. In terms of the occurrence of depression and diabetes together, studies suggest the prevalence of depression can be around 10-20% among patients with diabetes [8]. Previous studies have found that the odds of depression among individuals with diabetes can be 1.8-2.2 higher than that of the general population [9]. Unfortunately, the problem of undertreatment for depression is common to both individuals with and without diabetes. The omission of depression from the primary list of complications and comorbidities on the ADA website is partially indicative of the underdiagnosed and undertreated nature of this comorbid condition with diabetes. Therefore, these prevalence estimates probably underestimate the true prevalence of depression among individuals with diabetes [10]. Within the literature, there is significant variation in the reported prevalence of depression among this patient population, as the prevalence range uncovered in previous literature reviews spanned from 21.8-60.0% in controlled studies depending on depression measurement method [11-13].

Both depression and diabetes alone are challenging conditions for patients and providers to treat, thus it is almost inevitable that the combination of these two conditions compounds this difficulty [10]. This patient population is characterized by even worse health outcomes and greater health care costs than individuals who just have one disease or the other [14-16]. Consequently, the development of effective methods to improve the health of these patients in the long-term could have enduring benefits. In the past couple of years, researchers have identified a collaborative care treatment approach as beneficial for the treatment of depression among patients with diabetes [17]. These clinical trials have demonstrated positive effects on the depression status of these individuals in one or two years of follow-up [18]. Unfortunately, as is often the case with clinical trials, these findings are less robust due to the limitations of trial design. The lack of a generalizable patient population and short follow-up length limit the widespread applicability of this evidence. Accordingly, reviews of the literature call for a modeling-based approach to help fill in the evidence gaps, as better information is needed to convince health care providers such an intervention is worthwhile [18].

Thus, we develop a depression-prediction model within the context of diabetes and subsequently implement this prediction model into a diabetes microsimulation model. By developing a clinical depression prediction model for patients with diabetes and building it into an existing micro-simulation model of diabetes, we can use decision-modeling to understand and research the coexistence of depression and diabetes. Since depression is underdiagnosed and a very complex disease, there will remain a portion of the population that is missed by any type of prediction algorithm. Nevertheless, with a simple yet discriminating population model, we should be able to improve the clinical accuracy of a diabetes simulation model. Furthermore, we hope that researchers are able to continue modeling the interaction of depression with diabetes

and study the effectiveness of new treatments as they emerge. The following sections outline the epidemiology of diabetes, depression, and the two conditions in conjunction with one another. After this background information, the methodology for model development and the results are discussed. The second part of this chapter details implementation of the prediction model into the Michigan Model for Diabetes.

Diabetes

In general, diabetes is a disorder of glucose metabolism that results when the body cannot properly take up glucose for energy. During normal digestion, the body converts carbohydrates into glucose, a sugar, so it can enter the bloodstream. In normal human physiology, the presence of glucose in the bloodstream triggers the release of the hormone insulin from the pancreas. Insulin is one of the main metabolic control hormones in the body, and helps trigger cellular uptake of glucose as well as storage of glucose. For people with diabetes, the primary physiological defect is the lack of a proper insulin response to elevated glucose levels in the blood [5]. Type 1 diabetes is defined by a lack of insulin production by the pancreas, while Type 2 is defined by defective secretion of insulin and the development of insulin resistance, where cells stop responding to the increased insulin levels. The lack of insulin or a lack of a cellular response to insulin leads to increased levels of sugar in the bloodstream. Sustained hyperglycemia (high blood sugar levels) results in many negative consequences, including damage to blood vessels and nerves [5].

Due to the destructive effects elevated blood sugar levels can have on blood vessels and nerves, there are many associated complications and comorbidities with diabetes. Nephropathy is defined as malfunctioning of the kidneys. Over time with elevated blood sugar levels, there can be capillary damage, scarring, and cell growth to the kidneys, decreasing the ability of the kidney

to filter the blood to remove waste from the body. Damaged kidneys cannot filter as well, and gradually more protein makes its way into the urine [19]. Similar processes underlie the relationship between diabetes and the other macro- as well as micro-vascular conditions – stroke, cardiovascular disease, retinopathy, and neuropathy. Up to 97% of individuals with diabetes are dyslipidemic, where their blood is characterized by increased triglycerides and LDL cholesterol levels, as well as decreased HDL cholesterol levels. These characteristics are highly correlated with atherosclerosis, which leads to macrovascular problems. Furthermore, patients with diabetes have impaired regulation of blood flow, hyper-constricted blood vessels, and impaired circulation of nutrients and metabolic products between the blood and tissues, leading to the microvascular problems seen in nephropathy, neuropathy, and retinopathy. While there are many more mechanisms that associate diabetes with its many complications and comorbidities, this brief overview covers some of the commonly cited mechanisms while in-depth reviews of the other pathways have been published elsewhere [20].

Depression

Depression is generally thought to develop as a result of biological and social factors, from genetics and neurotransmitter levels, to environmental and psychosocial triggers. While there are multiple forms of depression, the most commonly referenced disorder when individuals say “depression” is major depressive disorder (MDD). An episode of major depressive disorder consists of decreased ability to work, sleep, eat, study, and enjoy life. This potentially life-threatening disorder can occur at any age, and its multiple etiologies increase the complexity of treatment [21]. Research suggests that 30-40% of the variance in susceptibility to MDD is influenced by genetic factors, while the remaining 60-70% can be attributable to individual-specific attributes. Stress sensitivity has been identified as one of many potential mechanisms

leading to depression, as increased stress levels increase the release of corticotropin-releasing-hormone (CRH) from the hypothalamus, which leads to elevated levels of cortisol in the blood. Studies have also shown that elevated CRH levels can be pathogenic for some types of depression [22]. Neurotransmitters have also been found to play an important role in depression pathophysiology. Depleted serotonin levels, which can be caused by increased monoamine oxidase(MAO), have led to mood alteration, altered behaviors, and disruption of affective inhibitory procession. In addition to these pathways, many possible hypotheses are also explored in other review articles [22].

Undiagnosed depression leads to greater medical care utilization. For example, research has found that almost one-quarter of the top 10% of health care utilizers suffer from major depression [23]. One analysis found that only 40% of patients with depression were receiving any type of treatment [24]. Patients with diagnosed depression have significantly higher primary care, specialty care, inpatient, pharmacy, and laboratory costs than individuals who did not exhibit symptoms of depression [25]. Other research has found that 72.1% of individuals with MDD in their lifetime reported having another mental health disorder, along with many patients receiving inadequate control of their symptoms [26]. Problems of underdiagnosis and undertreatment are compounded when an individual with depression has another comorbid condition [27-29]. Further, problems with treatment compliance can reduce effectiveness of medications needed to improve any existing illnesses [30].

Diabetes and Depression

There is no clear directional relationship between diabetes and depression. Studies have documented mechanisms by which depression can predispose individuals to diabetes, while others have uncovered ways a diagnosis of diabetes can lead to the development of depressive

symptoms. One of the proposed mechanisms from diabetes to depression is the psychosocial burden of having a chronic disease. The awareness of having a chronic illness, along with the perceptions of individual disability could burden individuals, especially those with low social support systems [8]. Considering the high level of self-care required for diabetes, individuals have to be able to deal with the increased responsibilities that fall upon them, thus a variety of individual and societal-level factors could predispose an individual with diabetes to develop depression [31]. Researchers have also suggested a possible biochemical link from diabetes to depression via increased nervous system arousal [32]. The evidence for the psychosocial association between diabetes and depression comes from studies showing prevalence levels of depression among individuals with high fasting plasma levels of glucose being lower than those with a formal diagnosis of diabetes [33]. Conversely, a separate study documented an increased relative risk of depression for individuals with high fasting glucose levels compared to those with normal glucose levels. Both studies show an increase in the relative risk of depression when fasting glucose levels are elevated. Furthermore, it would also be expected that the diagnosis of many diabetic complications can also be associated with depression.

Turning to the other directional possibility, researchers have looked at the risk of developing diabetes once diagnosed with depression. Meta-analyses of such studies have documented a 37-60% increased risk of developing diabetes for those who have depression or high depressive symptomatology [34, 35]. One mechanism authors have discussed for this proposed direction is the impact of depression on behavioral factors predisposing individuals to developing diabetes. Depression can be associated with unhealthy behaviors, high caloric diets, lower levels of physical activity, and subsequent high body-mass indices, all significant risk factors for developing diabetes [36]. The frequency of engaging in healthy behaviors can suffer

due to lower levels of interest and pleasure, or increased fatigue [37]. Studies have shown that individuals with depressive symptoms have fewer days of exercise and healthy diets, as well as higher rates of smoking than non-depressed individuals [38].

Researchers have found the interaction of depression and diabetes to have a synergistic effect on the risk for poor health outcomes among patients [39]. The effects of depression on quality of life can be quite large, and when depression develops within a context of pre-existing diabetes, the quality of life effects can be more than additive. Possible mechanisms for the synergistic effect of depression and diabetes on poorer health outcomes include the associated decrease in medication management, maintenance of a healthy diet and lifestyle (exercise and smoking) [40]. Previous work has shown that worse depression severity is associated with significantly worse adherence to diet and oral hypoglycemic medication, which can lead to worse glucose control and thus a synergistic effect on health and quality of life [36]. One study showed higher rates of diabetes for individuals with depressive symptoms at baseline, even when controlling for individual characteristics, but the addition of lifestyle variables (smoking, healthy diet, exercise) eliminated the relationship, indicating that the relationship between diabetes and depression may be mediated through such factors [41].

A biochemical pathway operating via increased cortisol release when an individual is depressed also has implications for insulin release by the body [42]. Because of evidence supporting both directions of the relationship, other researchers have settled on constructing the interaction between diabetes and depression as bi-directional. This coexistence of depression and diabetes is costly from many perspectives and thus merits further investigation. While researchers have and continue to investigate ways to improve treatment of depression among patients with diabetes using clinical trials, the role for simulation models is expanding.

Diabetes Simulation Models

There are a variety of existing clinical simulation models that work to replicate the progression of diabetes. Every two years, the teams that have developed these models compete and compare their simulation models [43]. The Michigan Model for Diabetes (MMD) is a publicly available microsimulation discrete-time diabetes model that allows users to simulate the progression of Type 2 diabetes including the development of cardiovascular disease, cerebrovascular disease, nephropathy, neuropathy, and retinopathy, as well as death. This model aggregates direct medical costs as well as quality-of-life estimates from diabetes and any complications and comorbidities that each individual develops in each cycle (1 year) of the model for a population cohort. The costs included in the MMD only capture diabetes specific costs and the costs associated with the ongoing treatment of the complications and comorbidities included in the baseline model. A state-transition model, the microsimulation model updates the health states of each individual based on transition probabilities that vary based on individual characteristics. The cohort for each simulation is user-defined, and can be generated by inputting population distribution estimates (average/standard deviation of age, gender, duration of diabetes, etc.). Generally speaking, the MMD is made up of multiple nested sub-processes that simulate the parallel progression of each of the modeled comorbidities. The model structure is provided in the Appendix. Each nested sub-process has its own model structure and health states associated with the progression of the specific disease. For example, the stroke sub-model has three states: no cerebrovascular disease, survived stroke, and stroke death, along with one event state: stroke occurrence. Individuals move through the states depending on the transition probabilities built into the MMD. The transition probabilities are a function of individual characteristics including age, sex, race, smoking status, HbA1c, body-mass-index, systolic blood

pressure, lipid levels, and comorbidity treatment status [43]. Articles further detailing the model development, validation, statistical methods, software, as well as applications have been published before [45-48]. The MMD currently does not incorporate depression status into its simulated progression of diabetes. Since patients with both diabetes and depression have increased likelihoods of developing common comorbidities and complications, lower quality-of-life, and increased health care costs compared to patients with only diabetes, the incorporation of depression status could help improve the clinical utility of the MMD.

In this analysis, we develop a clinical prediction model for depression within a population of individuals with diabetes. Since the MMD simulates the progression of diabetes, we focus on modeling the development of depression conditional on an individual already having diabetes. While the bidirectionality of the relationship between depression and diabetes makes the development of diabetes within a population of depressed individuals equally interesting, we chose to work within the MMD so that we could build upon an existing validated model. An informative risk prediction model would help improve the clinical understanding of what may be significantly related to patients with diabetes developing depression. While both depression and diabetes are physiologically and clinically complex, estimates of what factors are associated with higher rates of depression could help improve the care received by these individuals through modeling studies by providing a better characterization of disease progression for these patients. Due to the complex multi-mechanistic pathophysiology of depression, previous prediction models of depression have varied widely in terms of their predictor variables.

Depression prediction models within other contexts provide information on potential predictor variables of interest for patients with diabetes. Among patients with rheumatoid arthritis, pain and fatigue scores, number of comorbidities, and duration of arthritis were some of

the more important predictors of self-reported depression [49]. For patients recently diagnosed with cancer, immediate scores of anxiety and depression, along with advanced disease status and lower levels of family support were highly related to anxiety and depression prevalence at 6 months [50]. In an analysis to predict a depressive episode at 1-year follow-up among adolescents, researchers found a well-performing 20-variable model that included many components of depression screening questionnaires ("Over the past week you were bothered by things that usually don't bother you"), along with age, body-mass index, feelings of social anxiety, and social cohesion. Most of the 20 variables were ones drawn from a depression screener, so the researchers were able to develop a well-discriminating model to predict depression [51]. This reliance of depression prediction models on mental health history is also seen within the context of diabetes.

In a recent study, researchers found the following six variables to be significant predictors of depression among patients with diabetes in Malaysia: sex, ethnicity, marital status, duration of diabetes, psychiatric illness in the family and alcohol consumption. The most significant predictor in their study sample was the history of psychiatric illness [52]. In Table 1 of their article, Fisher et al list the factors that the literature suggests affect the prevalence of depression among individuals with diabetes. Their list includes: age, sex, race, marital status, social class, employment status, number of comorbidities, diabetes type, disability level, pain levels, social support levels, anxiety, negative life events, self-rated health, adverse social circumstances, well-being, affect and quality-of-life, and illness intrusiveness [12]. In their analyses, education levels, high impact of diabetes, and financial stress were significant independent predictors of depression among individuals with diabetes of Latino or European-American descent, while spousal conflict was also significant for the European-American

patients. Age, sex, duration, number of comorbidities, glucose control levels, and BMI were all insignificant predictors for their sample of patients [12]. Finally, in a very recent analysis developing a clinical prediction model for depression among patients with diabetes, Jin et al found sex, diabetes self-care score, number of complications, previous diagnosis of MDD, number of ICD-9 diagnoses in the past 6 months, chronic pain, and self-rated health status to be the necessary predictor variables for their best performing model (area-under-receiver-operating-characteristics curve (AUROC) = 0.81). The data population for this study was from 2 clinical trials on underserved, predominantly Hispanic patients with diabetes in the Los Angeles County area [53].

As can be seen from a comparison of the three preceding studies, differences exist in terms of the significant predictor variables when developing a clinical model to predict the development of depression among a population of individuals with diabetes. One drawback to the existing literature is that the population samples from each of these studies are all very specific, so external generalizability is limited. All three studies draw from sample populations that most likely have variability in their experiences with both diabetes and depression. One study relies on a patient population in Malaysia, another using 187 European-Americans and Latino patients, and the last one used patient data from clinics in the Los Angeles County Department of Health Services. The development of depression for each of these patient populations can probably be driven in part by the context of their disease experience, which will vary across these populations. This limitation hinders external generalizability of these models. Secondly, two of the prediction models required knowledge of depression or psychiatric illness history among either the individual or their family. The recurrent nature of depression (50% will experience one or more additional episodes, 80% with a history of two episodes will experience

recurrence) makes any knowledge of depression history important in predicting depression development in the future [54]. While clinicians with knowledge of depression history should incorporate this into their assessment of depression risk among patients with diabetes, this information might not always be available to draw upon in assessing depression risk. Jin et al point out that some of their predictors are not readily available in clinical settings [53]. Finally, none of the aforementioned studies performed any type of external validation or calibration. These two steps help validate prediction models, as the population sets they draw from can skew coefficients and hinder external applicability. Without the validation and calibration of a prediction model with an external dataset, it is difficult to discern how much of their results are due to their study populations. The transportability of a prediction model to different populations or settings is limited when no external calibration or validation is performed [55]. As Steyerberg et al. point out, external validation and updating of the model with a completely independent dataset is the best way to evaluate model performance and improve generalizability [56]. The aforementioned studies have not done this, so the influence of their predictor variables in determining depression status is applicable only to their specific patient populations.

In this analysis, we develop a model predicting depression status using a sample population drawn from the Health and Retirement Study. We then use data from the National Health And Nutrition Examination Survey to externally validate and calibrate our model so we are more confident of our predictors when implemented into an existing micro-simulation model of diabetes. By developing this prediction model and building depression status into the MMD, we hope to improve simulation studies of diabetes progression. With the incorporation of depression into the diabetes simulation model, we will be better able to both predict the diabetes-

related outcomes for this patient population, as well as identify effective and cost-effective treatments.

Methods

Data Sources

We use data from the RAND Health and Retirement Study (HRS) longitudinal dataset. The Health and Retirement Survey is a longitudinal panel study of approximately 20,000 elderly individuals in the United States, sponsored by the National Institute on Aging (grant number NIA U01AG009740) and is conducted by the University of Michigan. Surveys are administered every two years, and collect information related to income, work, assets, pension plans, health insurance, disability, physical health and functioning, cognitive functioning, and health care expenditures. Data from this study have been used widely in research studies, and have been used many times to study diabetes within a representative patient population [57-60]. The RAND HRS dataset compiles the data from the different HRS waves in an easy to use panel format, and tracks answers to demographic, health, financial and housing wealth, income, social security, pension, health insurance, family structure, retirement plans, and employment history, from survey respondents over time. We draw our analytic sample using data from 2000-2012 of the RAND HRS dataset. The advantages of this dataset include the consistent tracking of disease status in every survey, as well as the rich information regarding current as well as past medical history. The survey is nationally representative of the older adult population in the United States; the consistency of data collection and the inclusion of data pertaining to both depression and diabetes make this dataset very useful from a research perspective.

The HRS is a national survey of individuals over the age of 50, along with their spouses. There are six cohorts that have been surveyed over the last 24 years. The health information that is maintained in the RAND compiled dataset includes information on medical utilization and expenditures, self-rated health, along with condition-specific health history. The list of conditions includes hypertension, diabetes, cancer, lung disease, heart disease, stroke, psychiatric illnesses, arthritis, back pain, and ulcers. Since diabetes is not the main focus of this survey, there are many characteristics related specifically to diabetes and diabetes management that are not captured. For example, HbA1c levels or a full history of diabetic complications is not available from the RAND dataset. While the Health and Retirement Survey tracks depression over time, because of the breadth of the survey and inherent time constraints, an abridged depression questionnaire is used instead of the more common full set of questions. Nevertheless, with this dataset, we are able to track many individuals over time with regards to their health status. Furthermore, the sample size is large so we are able to generate estimates from a much larger sample than is normally done in similar analyses.

Depression Status Variable

The HRS survey began to ask respondents a shortened version of the Center for Epidemiological Studies-Depression (CES-D) questionnaire in their first wave, though this went through some changes after wave 1. From 1994 onwards, respondents have been asked an eight-question survey that provides a measure of depressive symptomatology (Table 3.1). The original 20-item CES-D scale helps measure frequency of depression symptoms. Answers to the survey are summed, ranging from 0-20. While the original intent was not to track the prevalence of depression, many previous studies have done so and found ranges of sensitivity of the full CES-D from 70%-99%, and ranges of specificity from 56%-94% [61]. The original cutoff point for

indicating depression likelihood corresponded to the 80th percentile of the CES-D 20-point scale. Due to time constraints of the HRS survey, the eight item question was more feasible (Table 3.1). To replicate how the original cutoff point was determined for the likelihood of depression, we calculated the 80th percentile of the distribution of responses in the whole dataset. This value was then confirmed by comparing it with the 80th percentile of the distribution of scores from each individual wave. Figure 3.1 shows the distribution of the CES-D scores from 2000-2012. We determine the cutoff point for depression status from the overall population without restricting to the diabetes-specific population as we did not want the prevalence of diabetes to artificially increase the threshold. The 80th percentile of this distribution corresponds to a score of 3 on the 8-point scale, and is the cutoff we use to distinguish depression vs non-depression throughout our analysis.

Table 3.1 - Health and Retirement Survey 8-item CES-D Scale

Much of the time during the past week…(Yes/No?)

1) You felt depressed
2) You felt that everything you did was an effort
3) Your sleep was restless
4) You were happy (reverse-scored)
5) You felt lonely
6) You enjoyed life (reverse-scored)
7) You felt sad
8) You could not get going

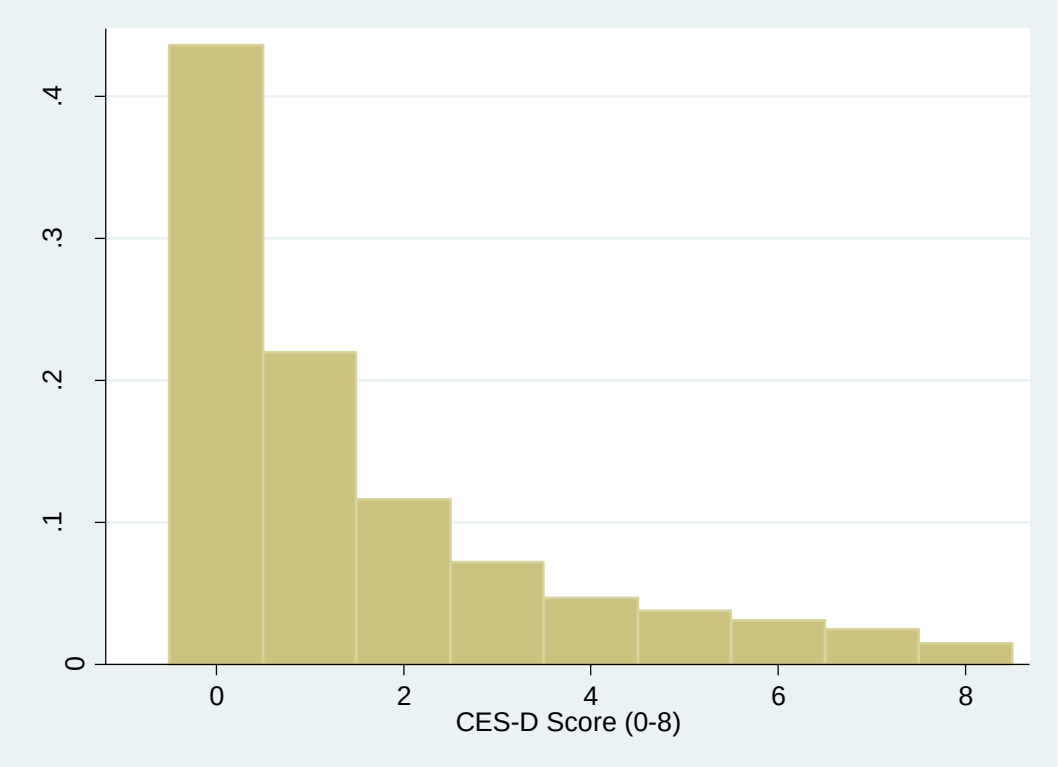


Figure 3.1 - Distribution of CES-D Scores for all respondents 2000-2012

All individuals who have CES-D scores greater than or equal to 3 are coded as having depression for our analysis. For example, of all HRS respondents in 2000 with a CES-D score, 24% would be coded as having depression. Among individuals with diabetes in 2000, 34% would be coded as having depression when using the same cutoff of 3. This provides us with the dependent variable of interest, as we want to determine how individual demographic and health characteristics may or may not predict the development of depression.

Sample

As we are interested in modeling the development of depression conditional upon an individual having diabetes, we restrict our analyses to the subset of participants that report having clinical diabetes. This is tracked in each wave of HRS, so all participants who report being diagnosed with diabetes are kept in our study sample. Our identification strategy relies on individuals who report having diabetes going from a non-depressed state to a depressed state, so

we must also limit our sample to those individuals who do not have depression at baseline. If we did not have this exclusion criterion, then it would be difficult to parse out what characteristics were determinants of depression development due to left-censoring of the data. Instead, with this exclusion, we are able to use regression analysis to identify variables that are significant in the development of depression. Our exclusion of individuals leads to a final sample size of 1,749 individuals with data over 7 waves (seen in Figure 3.2). Of the 3,010 individuals who had diabetes in 2000, 888 of them had a CES-D score greater than or equal to 3, 367 of them did not have CES-D scores, and 6 of them did not have complete data for the independent variables of interest.

Figure 3.2 - Sample Development

Independent Variables

As mentioned before, the literature suggests that significant predictors of major depressive disorder include sex, age, education level, marital status, number of complaints, and a variety of characteristics related to depression symptoms [62]. Another recent study developed a clinical forecasting model for depression among patients with diabetes using a machine learning approach and found sex, number of complications, history of depression, and number of other comorbidities to be significant predictors of depression status [53]. The RAND HRS data allow us to track each participant's health status across a variety of conditions, including: lung disease,

heart disease, stroke, cancer, arthritis, hypertension, and a broad psychiatric illness category. The HRS survey collects data on both current disease experiences as well as diseases that individuals had in the past. Additionally, the dataset also provides information related to their age, sex, race/ethnicity, body-mass index (BMI), and years of education. Table 3.2 shows descriptive statistics for the analytic sample across these variables of interest. Since we would expect there to be differences in the relationship between age and depression likelihood, we plot the proportion of individuals with depression across the spectrum of ages in our analytic sample (Figure 3.3). The relationship between depression and years of education and age slopes in the direction that matches previous literature, providing some more validity to our measure of depression [63]. The lines in these graphs are from a locally weighted smoothing function. While Fisher et al detailed a list of all variables that correlate with depression among patients with diabetes, including all of these as predictor variables would be impractical and unwieldy, and unavailable from our data sources.[12] We use this comprehensive set of potential predictor variables to narrow down the variables of interest from the RAND HRS dataset. These variables fall into two categories – demographic and health care information. Then, we had to consider the variables that the existing simulation model of diabetes included, as we would not be able to use any predictor variables if the simulation model did not have existing parameters for them. The Appendix provides a flowchart detailing our variable selection process.

Model Development

Overall, the steps of clinical prediction model development include sample selection, dependent variable definition, predictor variable set determination, regression analysis to determine predictor coefficients, model performance comparisons, and then validation and calibration with an external dataset [64]. The process used to develop our prediction model for

depression within the context of diabetes was an iterative one, where different models with varying predictor variables were used to find what combination resulted in the best model performance. We used the area under the receiver operating characteristics curve (AUROC) to compare model performance. We show the final set of models we estimated after exploring many less informative models (Table 3.3). Each specification used a random effects panel logistic regression with depression status as the dependent variable.

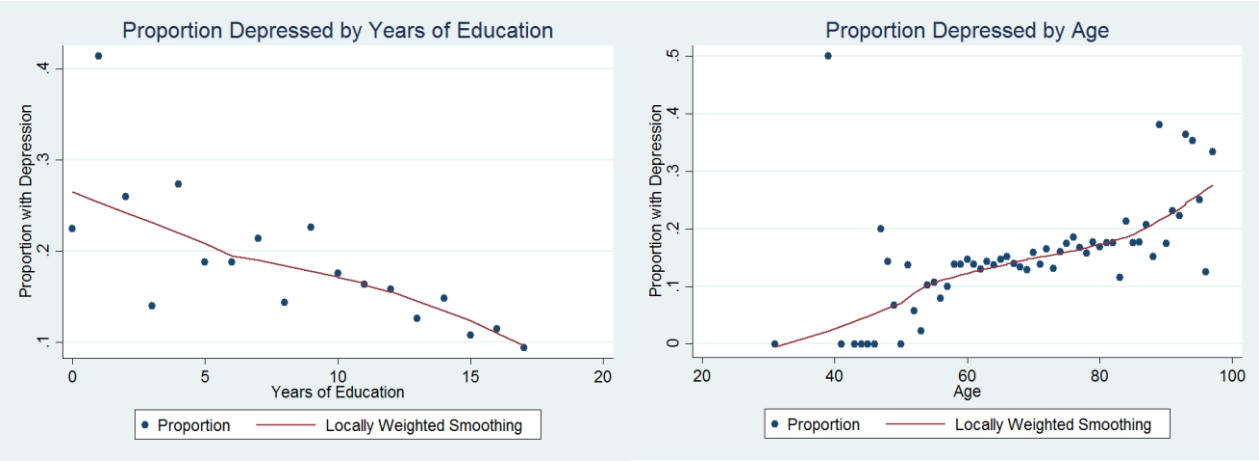


Figure 3.3 - Relationship between Depression and Age

We used a random effects specification instead of fixed effects because we were interested in capturing sex effects on depression development as the existing research suggests that females with diabetes have a higher prevalence of depression than males with diabetes [65].

In the first specification, only basic demographic variables were included – age, sex, and race. Then, with each additional specification, we added the current BMI level as well as health status indicator variables for patient history of hypertension, stroke, and heart disease. We ran separate models with more detailed health information, but the Akaike Information Criteria (AIC) increased with the addition of these largely unrelated variables. Once all the pertinent

health variables had been added, we included a time variable to see what the effect of each additional year with diabetes had on the likelihood of developing depression.

Table 3.2 - HRS Population Characteristics (2000-2012)

Gender	Mean
Male	49%
Female	51%
Age (in 2000)	67.8
Race	
Caucasian	76%
African-American	20%
Other	4%
BMI (kg/m^2)	29.7
Education Category	
Less Than High School	31%
GED	5%
HS Graduate	30%
Some College	19%
College and Above	15%
History of…	
Hypertension	75%
Heart Disease	35%
Stroke	14%
Lung Disease	8%
Arthritis	64%
Cancer	17%
Psychiatric Illness	12%
Sample Size	1,749

The duration of diabetes variable from HRS has a lot of missing data, so instead of relying on imputed data points, we include a "time" variable that we conceptualize as a left-truncated proxy for duration of diabetes. While we do not know the exact year of diabetes

diagnosis for the 1,749 individuals in our analytic sample, we know that within the timeframe of our analysis, their duration of diabetes increases. So this time variable captures this increase in duration. This was the full specification, but then as the Michigan Model for Diabetes is limited in terms of which variables are tracked over time, we had to then restrict the full specification to what was feasible within the current Michigan Model (the Appendix shows the variable selection process).

Table 3.3 - Model Specification Process

Model	Predictor Variables
1	Age, Race, Sex
2	1 + BMI, Smoking Status, Hypertension Indicator
3	2 + Heart Disease in Past 2 Years
4	3 + Stroke in Past 2 Years
5	4 + History of Heart Disease or Stroke
6	5 + Time
7	Age, Race, Sex, BMI, Smoking Status, Hypertension Indicator, History of Stroke, History of Heart Disease, Time
8	7 + Age as a Piecewise Function
9	Removed Age from Model

The differences in the AUROC statistics for each of these specifications provide us with information about model performance (Figure 3.4). Model specifications 6 and 8 are the ones with the highest model performance, and Model 8 is the model which provides us with the best discrimination and maintains feasibility of implementation within the existing simulation model of diabetes progression.

The variables that are most significant in predicting depression status among patients with diabetes are sex (female), BMI, hypertension, history of stroke, history of heart disease, and

time. We then tested the significance of groups of related variables by evaluating the chi-square test results, Akaike Information Criteria (AIC), and AUROC statistics from iterations of Model 8 and 9 with and without variables. The variables that did not have a significant effect on the overall fit of the model were dropped (race, smoking status, and age). Although we were not concerned with overfitting our model, adding variables that have little added value is inherently unnecessary for a clinical prediction model that needs to be simple to maintain practicality. Since age increased the AIC of our model, we dropped it from the model.

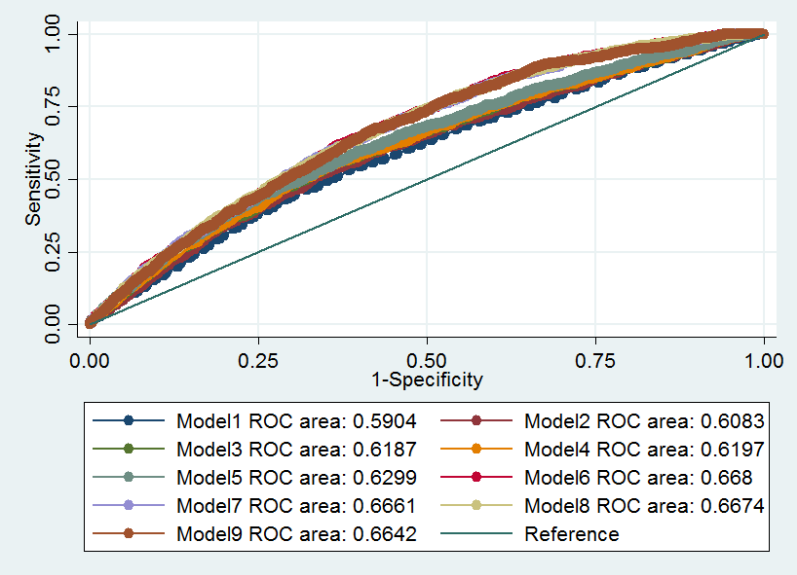


Figure 3.4 - ROC Curves Across All Models

The resulting model, Model 9 in Figure 3.4, has the following independent variables: sex, BMI, hypertension, history of stroke, history of heart disease, and time. Males had a large, significantly negative effect on the likelihood of being depressed compared to females. This is similar to what other studies have shown, with higher rates of depression among females with diabetes compared to males with diabetes. The next two largest coefficients were significantly positive, as a history of stroke and heart disease both increased the likelihood of depression.

Again, this matches the literature as number of complications/comorbidities and health status have been shown to be related to depression development. Current hypertension, time, and BMI were also significantly positive at different significance levels. Although hypertension specifically has not shown up in the literature as a significant predictor before, it may have been included in measures of self-rated health as well as numbers of complications and comorbidities. Time, which we view as a proxy for duration, has been mentioned multiple times in other studies as an important possible determinant of depression development. Finally, BMI is also a measure one that has been used in depression prediction models before (adolescents), and may be construed as a measure of health status. We then calculate a predicted probability of the dependent variable being equal to 1 for each individual using the following equation:

$$Pr\ (y_{it} = 1) = \frac{exp(X_{it}\beta + \mu_i)}{1 + exp(X_{it}\beta + \mu_i)} \qquad \textit{where we assume that } \mu_i = 0$$

Then, to dichotomize each individual as either having depression or being depression-free, we use a predicted probability cut-point where everyone with a predicted probability above the cutpoint is assigned a positive depression status. Applying our prediction model to the RAND HRS data, we were able to calculate the sensitivity and specificity of the algorithm across a variety of different cut-points. The sensitivity and specificity of our model using an optimal cutoff resulted in a sensitivity of 67%, and a specificity of 58%. This cutoff corresponded with the maximum AUROC of 0.66. It is important to note that these results are specific to the RAND HRS population that we rely on to develop our model.

Robustness Checks

One of the main variables of interest in predicting depression development is the duration of diabetes. As this disease is one where the development of comorbidities and complications

occurs gradually, the duration of diabetes can be a strong predictor of the development of other illnesses. Accordingly, we were interested in determining whether or not the duration of diabetes is a significant predictor of depression status in our dataset. Since the duration variable in HRS has incomplete data, we decided to perform a robustness analysis with our own constructed measure of duration. For this, we limited our sample to individuals who do not have diabetes at baseline in our sample, but do develop diabetes at some point between 2000 and 2012. Then, for those individuals who do develop diabetes, we run a logistic regression with the same health and demographic variables as in our final model, but include duration of diabetes as a predictor (β=0.023, p=0.551; overall model c-statistic = 0.65). For the sample of individuals who develop diabetes, the duration of this illness is not a significant predictor, and does not improve model discrimination. One plausible explanation for this is that the average duration of diabetes when we restrict it to individuals who newly develop diabetes is only 6 years on average, so the dataset does not follow these individuals long enough for us to capture the effects of duration on complication and depression development after many years. Accordingly, we believe the "time" variable captures a censored proxy for duration, and with calibration of the coefficient using another dataset, we will have a reliable input for our prediction model. Another possibility is that duration is inherently collinear with other variables capturing complication history. Therefore, this relationship will need to be explored in further research with more in-depth data on individuals over time.

We were also interested in exploring the effects of age vs time in our model, as there is collinearity between age and time. To explore the effects of each of these variables in isolation, we first specified a model that excluded age, and then ran another model that excluded time. Comparing the c-statistics of these revised models provides estimates of the additional benefit of

including these variables in our model. Removing age from the model and only including time results in a small drop in our ROC statistic (c = 0.6656), while if we remove time from the model and only include age, our ROC statistic drops even more (c = 0.6287). The AIC's improved when we dropped age, but not when time was taken out of the model, so we decided to drop age and keep time in the final model. We then make the assumption that the time variable is a proxy for duration, and the coefficient estimate can be calibrated after applying to another dataset with better duration information.

External Validity

We ran external validity tests on our final depression prediction model using the National Health and Nutrition Examination (NHANES) cross-sectional survey from 2007-2008. This survey is used to assess the health and nutrition status of individuals in the United States, and provides information regarding depression status, current health status, and medical conditions status (including diabetes), in addition to a variety of other individual information. Compared to the HRS population, this survey population is younger (average age = 61.5 years), more diverse (30% not Caucasian or African-American), slightly heavier (average BMI = 32.4 kg/m^2), and slightly healthier (40% with hypertension). NHANES also collects duration information for diseases, so we know the first year of a diabetes diagnosis. Accordingly, we use this duration variable as our "time" input into our model. As mentioned earlier, validation and parameter updating of a prediction model with a fully independent dataset helps improve model generalizability.

To account for these differences in the sample populations, we manually calibrate the constant and coefficient estimates. Manual calibration or manual search is an iterative process where you alter input parameters to optimize outcome targets. In our analysis, the input

parameters are our predictor variable coefficients, and our targets are model performance which are specified as sensitivity, specificity, and the AUROC. Since the coefficients from our regression model is specific to the primary dataset, calibration of the parameters to optimize model performance in another dataset improves the reliability of the prediction model. As there are BMI and diabetes duration differences (not left censored in NHANES) between the NHANES dataset and the HRS data, we calibrate these coefficients to improve model performance. Finally, we calibrated the other coefficients to improve model performance. This calibration process was an iterative one where we cycled through changes to the coefficients and evaluated the sensitivity, specificity, and AUROC of the prediction model. After extensive manual calibration of these coefficients, a final model with a predicted probability cutoff of 0.5013 resulted in a AUROC of 0.59, a sensitivity of 61% and a specificity of 54% (coefficients seen in Table 3.4). These statistics are similar to what has been shown in the past, as another prediction model of depression among patients with diabetes had a sensitivity of 50% when applied to a different sample population [52].

We use calibrated coefficients because the NHANES dataset is more nationally representative, so we believe it will reduce the bias in our predictions when applied to the MMD compared to using the uncalibrated model from the HRS dataset. After calibration, the magnitude of the coefficient for the duration variable decreased, corresponding with our assumption that the original time variable from the RAND HRS was a left-truncated proxy for duration. In the original sample, the duration of diabetes would be higher than the time variable for any individuals who had been diagnosed with diabetes before 200. Through the external validation and calibration, our coefficient estimate was reduced to a more realistic point estimate since the duration variable will increase for every individual in every cycle while the other

variables do not have this guaranteed monotonic increase over time. Finally, we calibrated the coefficients on sex and health status to improve model performance. Initial calibration of the model of only the BMI, duration, and constants led to a disproportionate share of depressed individuals being female (98% in year 10). As we want to implement this prediction model using a more representative population of patients with diabetes, we needed to make sure that our depression prediction algorithm did not have an extremely unrealistic sex distribution.

Table 3.4 - Calibrated Regression Coefficients for Depression Prediction Model in Patients with Diabetes

Predictor	Coefficient
Male	-0.262358
BMI	0.0158003
Hypertension	0.3862508
History of Stroke	0.5899783
History of Heart Disease	0.6835347
Duration	0.0120267
Constant	-0.845155

Thus, we increased the coefficient on male, and slightly increased the coefficients on the health status variables. To evaluate performance of this updated model, we applied the coefficients to both the RAND HRS dataset as well as the NHANES dataset. Figure 3.6 shows the two ROC curves for the final model when applied to these samples. In the RAND HRS sample, applying the final model to the sample resulted in a sensitivity of 62%, specificity of 56%, and a AUROC of 0.62, compared to a sensitivity of 67%, specificity of 58%, and a

AUROC of 0.66 in the original model. Using the same prediction model in the NHANES dataset resulted in a sensitivity of 61%, a specificity of 54%, and a AUROC of 0.59. The slightly better performance in the RAND HRS sample is expected since the foundation for the coefficients was from this dataset.

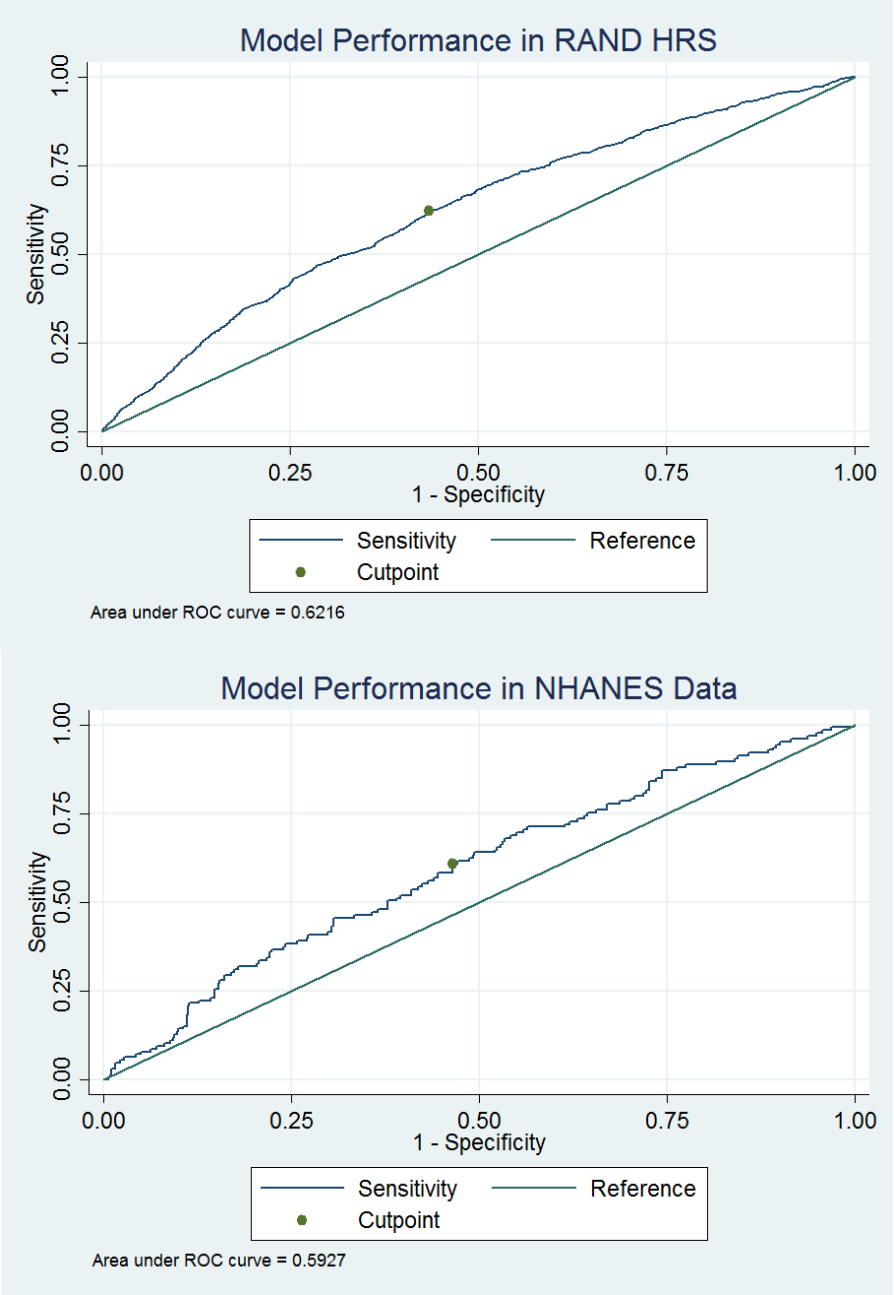


Figure 3.5 - Model Performance

Limitations

Any modeling study must rely on assumptions. One important one that we make in our analysis is around the logistic regression approach. In our data, depression is coded as a binary indicator variable. For our purposes, we used a panel random effects logistic regression model to evaluate the coefficients on the predictor variables of interest. First, we chose a random effects approach rather than a fixed effects model because we were interested in the effect of many variables that are time-invariant (sex and potentially disease history). In a fixed-effects specification, these variables would have been dropped from the model, and thus we would have been unable to estimate their effects on depression development. With the random-effects model, we assume that the variation across individuals is random and uncorrelated with the other independent variables. Unfortunately, the random effect is not actually estimated in a panel logistic regression, so to calculate predicted probabilities, we must assume that the individual-level random effect is 0. So we have to make an assumption that the random effect is 0, while the assumption that there is variation in the random effect is the underlying justification for the random effect model. We believe that this simplification of the random effects being 0 is reasonable because this assumption impacts the predicted probabilities from our regression results. In the step following initial regression analysis from the RAND HRS data, we calibrated our coefficients to the NHANES dataset. The calibration procedure involved changing the constant as well as the coefficients on other variables. Since we change our coefficients for model calibration, the assumption of the random-effects being 0 plays a less important role in driving the results since this assumption primarily biases the coefficients. The other option for model specification was a linear probability model, but the possibility of predicted probabilities outside the range of 0 and 1 rendered this approach less useful. We ran our analyses using a

linear probability model and it resulted in numerous out of range predicted probabilities. Thus we decided to use the logistic model for our purposes.

From a data standpoint, the 8-item CES-D questionnaire is definitely less precise in identifying clinical depression than the longer questionnaire or other depressive symptom scales. The literature suggests that the CES-D summary score can be used to dichotomize depression status, and the cutoff of 3 has been found to be associated with a sensitivity of 71% and specificity of 79% [61]. Another limitation is that our dataset for external validation and calibration is cross-sectional, while the model estimates are from a panel dataset. It would have been more ideal to calibrate the model with another panel dataset. Since our model will be implemented in a microsimulation model that simulates a cohort population over time, calibration and validation with two panel datasets would have increased the reliability of the estimates. Nevertheless, data availability limitations and the relative scarcity of nationally representative panel datasets compared to cross-sectional data necessitate the usage of NHANES as our secondary dataset. Finally, this prediction model operates under the assumption of diabetes leading to depression, and not the reverse direction. We do not evaluate what variables are important in predicting the development of diabetes among depressed individuals. Since both directions have viable mechanisms, it is equally important to understand what may predispose individuals with depression to develop diabetes. As the clinical prediction model is constructed under the context of a diabetes simulation model, we focus only on the development of depression for patients with pre-existing diabetes. The developed model still improves our understanding of the interaction between these two complex diseases.

The relationship between duration of diabetes and risk of depression is most likely complex and non-linear. Research has shown that the association between duration of diabetes

and depression may be “J-shaped”, with an increase right after diagnosis, followed by a drop in odds of depression, and then an increase over time [66]. This non-linear relationship merits further investigation. Unfortunately, the data availability issues of the RAND HRS dataset which we use for predictor variable coefficient determination prevent us from exploring any potential non-linear effects of duration on depression risk among a population of patients with diabetes. Future work with more robust data could better inform the non-linearities that may exist between the duration of diabetes and depression risk to improve model performance. The Health and Retirement Study focuses on the elderly population, as the average age of the analytic sample in the first year of our dataset is 67.8. The average age of the population from our simulations is approximately 53. Although we calibrate the prediction model with NHANES data where the population is younger than the HRS population, we cannot guarantee that calibration has fully accounted for the differences in the age distribution between the population of interest and the data sources. Therefore, our model may inaccurately identify younger individuals as not having depression. This would bias our results by reducing the estimates of the benefits of treatment, as the successful treatment of younger individuals with depression would lead to more years of healthy life and more averted health care costs by delaying the development of diabetic complications and comorbidities. Any analyses evaluating the benefits of treating individuals with diabetes using our modified model would not capture all the potential benefits of treatment, especially within this younger population.

Finally, our prediction model is not the most discriminating algorithm, as it results in sensitivity ranges from 61-62%. Many of the aforementioned work evaluating the development of depression among patients with diabetes have relied on knowledge of mental health history as a predictor variable. Without this knowledge, it inevitably will become more difficult to discern

independent effects of individual characteristics that are associated with increased likelihoods of developing depression. Nevertheless, future work will center around improving the depression prediction algorithm. As the United States Preventive Services Task Force Recommendation Statement in January of 2016 recommends depression screening of all adults 18 years and older, the identification of individuals at risk of depression should increase in the coming years [67]. With this potential for improvements in depression identification, better data may become available to develop a more discriminating prediction model.

Model Implementation

The previous section has detailed the necessary steps to assign individuals depression status within the Michigan Model for Diabetes. After an individual is assigned as being either depressed or not depressed, they will have to experience changes to their transition probabilities that dictate their diabetes progression over time. Previous work has documented the increased risk of developing the diabetic microvascular and macrovascular complications and comorbidities for individuals who are depressed [68]. We apply the increased risks of transitioning into each of the corresponding sub-states of the Michigan Model for Diabetes for depressed individuals by using the values derived from the Lin et al study (seen in Table 3.5) [68]. The numbers in Table 3.5 show the mean and standard deviations of the increases. For example, compared to a non-depressed diabetic patient, someone with both depression and diabetes will experience an approximately 25% higher risk of transitioning into the cardiovascular disease sub-model. We only apply these increases in transition probabilities to the first transition within each of the sub-states (transitions seen in Figure 3.7). The increases in transition probabilities reported in the literature are based on a comparison of patients with both depression and diabetes to patients with diabetes and no depression, while the transitions built

into the MMD are based on a population with and without depression. Since we separate individuals into a depressed or non-depressed characterization, we increase the transitions for patients with depression, and decrease the risk for patients without depression. To do so, we assume that the prevalence of depression among patients with diabetes is 20% [9], and adjust the transition probabilities using a weighted average modification (Appendix).

The Appendix shows each of the disease sub-processes that are modeled in the Michigan Model for Diabetes. In terms of implementation, we add functions that calculate the numerator and denominator of the predicted probability for every individual based on their characteristics. Then, we add a covariate that recalculates the predicted probability for each individual in each cycle. Finally, the depression covariate assigns an individual a 1 or a 0 for depression status based on the predicted probability being above the threshold. Based on each individual's assigned depression status, the model will then apply the changes to the transition probabilities.

Table 3.5 - Multiplicative Factors

Transition Probabilities	**Diabetes + Depression**
Neuropathy Subtree	1.36 (Normal, 1.36, .1775)
Retinopathy Subtree	1.36 (Normal, 1.36, .1775)
Nephropathy Subtree	1.36 (Normal, 1.36, .1775)
CVD Subtree	1.25(Normal, 1.25, .135)
CHD Subtree	1.25(Normal, 1.25, .135)
No Comorbidities	1

No nephropathy → microalbuminuria

No neuropathy → clinical neuropathy

No macular edema → macular edema or proliferative retinopathy

No cardiovascular disease → angina or congestive heart failure without myocardial infarction, coronary artery disease without myocardial infarction, myocardial infarction, death.

No stroke → stroke

Figure 3.6 - List of Transition Probabilities Modified

The Appendix provides the diagrams from the MMD Manual that show each of the sub-states that the MMD models as well as the overall model structure. Each of the initiating steps for the sub-models can be seen in the Appendix. There are multiple "initiating" steps for individuals in the retinopathy and cardiovascular disease sub-states as these processes have multiple possible disease stages that an individual can transition into from a healthy state, and each of these initiating steps have an increase in their transition probability when an individual is depressed. These changes allow us to model the development of depression among a patient population with diabetes as well as the subsequent changes in the risks of the common comorbidities and complications associated with diabetes.

Another important consideration with the implementation of a depression status variable in a model of diabetes progression is the effect of depression on mortality. Although depression can be a strong independent risk factor for mortality [69], in the context of diabetes the independence of this relationship becomes less apparent. Research into the risk of all-cause mortality among patients with both diabetes and depression has been mixed. After controlling for a variety of individual characteristics, including numbers of complications, one study found

major and minor depression to be significant predictors for mortality [70]. In another prospective study however, researchers found that after controlling for microvascular and macrovascular complications, depression was not significantly related to increases in all-cause mortality or cardiac mortality [71]. Therefore, since there is mixed evidence, we decide to allow the effect of depression on mortality to operate through increased risks of the microvascular and macrovascular complications. In the Michigan Model for Diabetes, individuals can reach the terminal death state from end-stage renal disease, cerebrovascular disease, coronary heart disease, and an alternative route to death (Appendix). The risk factors underlying the transition to mortality within some of the disease sub-models include age, gender, race, smoking status, hemoglobin A1c, systolic blood pressure, lipid levels, and medication usage. In others, a specified percentage of the population transition to death in each cycle. With the increased risk of progressing into the first state of each modeled complication and comorbidity, individuals with depression should then be at a higher risk of mortality than patients without depression. Therefore, we believe that our modifications to the MMD should result in individuals with depression experiencing a higher rate of complications and comorbidities, as well as mortality. If there is an independent risk of mortality for patients with depression, then our model would underestimate the calculated benefit of treating depression among patients with diabetes as treatment would avert more deaths compared to no increase in the treatment of patients with depression. As more research is done in this area, we will modify the model as necessary to improve clinical accuracy.

Previous descriptive work has documented the increases in health care costs for patients with diabetes and depression compared to patients with diabetes alone. These cost increases have ranged from 37-97% higher for patients with both depression and diabetes. Since the literature

does not identify the independent effect of depression on health care costs, these cost increases can be attributed to the increased risk of complications and comorbidities that individuals with both depression and diabetes face. Accordingly, we do not increase the estimates of annual health care costs for individuals based solely on their depression status. Instead, the increased transition probabilities into the different disease states in the MMD should lead to a natural increase in health care costs. This way, we avoid double-counting health care cost estimates, and develop conservative estimates of the health care costs associated with having depression and diabetes.

Conversely, there is a great deal of literature surrounding the utility decrements that exist for individuals who are depressed, even when controlling for their other disease states. In hierarchical regression analyses controlling for complications, medical history, and demographic information, researchers found a negative and significant impact of depressive symptoms on measures of quality of life [72]. Others have also found significantly lower scores on quality of life scales for patients with both depression and diabetes when controlling for some patient characteristics [73, 74]. Thus, we believe that there is a negative effect of having depression on an individual's health-related quality of life that operates beyond the increased risks of developing complications and comorbidities. To properly incorporate this negative decrement into the calculated annual quality-of-life scores, we decrease this estimate by 0.10 if an individual is assigned a positive depression status in that year [75]. There is very limited literature on the raw independent effect of depression on the Quality of Well-Being (QWB) scale, which is the scoring mechanism used in the MMD. The aforementioned literature provides coefficients of effects on other common quality of life scales (SF-12 or SF-36), but since these scales have different baseline values and scoring mechanisms, these coefficients cannot be

translated into a decrement that can be used for the MMD and its QWB scale. The literature surrounding any quality impacts of depression captured via the QWB scale are from over 20 years ago, but provide the basis for our 0.10 decrement. In a longitudinal cohort study of adults in a community population, Fryback et al. documented a 0.08 difference for individuals with depression vs. those without depression [76]. In another study, differences between non-depressed and depressed patients ranged from 0.14 and higher, while the difference between patients exhibiting severe depressive symptoms vs mild depressive symptoms was between 0.10 and 0.12 [77, 78]. The baseline QWB score in the MMD is drawn from work by Coffey et al as well as Zhang et al, but both these studies do not provide data on the incremental effect of depression [79, 80]. Therefore, we apply a 0.10 utility decrement for individuals with depression, with a range of 0.05-0.16 in sensitivity analyses.

To avoid making the depression characteristic an unrealistic absorbing state, we also implement a usual care treatment parameter. Usual care consists of standard care by primary care physicians, with no extra emphasis on treating depression. Physicians can administer pharmacotherapy and refer patients to psychiatrists. Studies suggest that the usual care treatment approach can be successful in approximately 30-40% of individuals who have depression [81-83]. There are no precise ways to predict which individuals will respond to treatment, so we randomly assign successful treatment to individuals using a Bernoulli distribution. So in each year, of the population with depression, approximately 40% should experience successful treatment. The cost of administering the usual care treatment is approximately $402 [84]. We update all costs in our analyses to March 2016 US Dollars by using the Consumer Price Index.

To summarize, we used data from the RAND Health and Retirement Survey as well as the National Health and Nutrition Examination Survey to build a model that calculates the

predicted probability of an individual with diabetes being depressed in a given year. Our predictor variables were female gender, body-mass index, current hypertension, history of stroke, history of heart disease, and duration of diabetes. We then took this prediction model and incorporated it into the Michigan Model for Diabetes, an existing microsimulation model of diabetes progression. Predicted depression status in the microsimulation model at the beginning of each cycle determines if an individual will experience increases in their probabilities of initiating into each of the sub-disease processes currently included in the microsimulation model (neuropathy, nephropathy, retinopathy, cardiovascular disease, and cerebrovascular disease). These increases in comorbidity and complication development should lead to higher direct medical costs on average, so we do not add any extra depression-associated costs. On the other hand, as depression foreseeably should independently reduce individual quality of life, we include a utility decrement that takes place whenever someone is assigned a positive depression status. We also incorporate a usual care treatment parameter so that a percentage of patients with predicted depression are successfully treated in each cycle. These modifications to the Michigan Model for Diabetes should allow research into the progression of diabetes with depression, as well as treatment possibilities for this patient population.

Simulation Results

In this section, we provide data from a simulation using our modifications to the Michigan Model for Diabetes. We simulated 10,000 individuals with diabetes for 20 years, and all costs are reported in 2016 US Dollars using the Consumer Price Index to account for inflation, and a discount rate of 3% is used. The cohort is drawn from a distribution of population inputs, where the average age is 53, with a standard deviation of 7 years. The average duration of diabetes of the population is 5 years. The MMD generates a simulation cohort using the

population input distributions. As seen in period 1, 17.2% of the cohort is depressed (Table 3.6). This prevalence is within the ranges estimated in the literature. Some differences exist in period 1 between depressed and non-depressed individuals in terms of disease status. A significantly greater proportion of the depressed patient population is in one of the cardiovascular disease health states than the non-depressed population (76% in the No CVD state vs. 94% in the No CVD state). For our simulation tables, they are grouped by the disease process in the leftmost column. In the other disease sub-states, the differences between the depressed and non-depressed patients is less borne out in the first period of the simulation. While significantly more depressed patients experienced strokes, there were smaller differences for these patients in the no neuropathy or no nephropathy states. The estimated health care costs and utility scores are significantly different for depressed individuals in period 1 though, with the health care costs being 39% higher. These results align with the findings in the literature discussed earlier.

Table 3.6 - Simulation Results (1000 individuals over 20 years)

	Period 1		Period 10		Period 20	
Characteristics	*Non Depressed*	*Depressed*	*Non Depressed*	*Depressed*	*Non Depressed*	*Depressed*
Count	8284	1716	6508	2536	4132	3128
Average Age	53.98	53.96	62.57	62.89	72.26	71.67
Male	75%	17%	68%	52%	67%	55%
Average BMI	31.6	31.8	32.9	33.4	34.7	35.3
Smoke	28%	29%	18%	12%	10%	8%
No Stroke	99%	94%	98%	94%	97%	94%
Survive Stroke	1%	6%	2%	6%	2%	5%
No CVD	94%	76%	84%	44%	71%	43%
Angina	2%	1%	4%	9%	4%	7%
CHF w/o AMI	1%	1%	5%	17%	12%	24%
CAD w/Proc	0%	0%	2%	7%	5%	8%
Survive MI	3%	21%	4%	19%	7%	13%
CHF	0%	0%	0%	2%	2%	4%
No Nephropathy	86%	85%	54%	51%	34%	30%
Microalbuminuria	7%	8%	23%	26%	20%	23%
Proteinuria	7%	7%	23%	23%	45%	45%
ESRD Dialysis	0%	0%	0%	0%	1%	1%
ESRD Transplant	0%	0%	0%	0%	0%	0%
No Neuropathy	86%	85%	53%	51%	32%	29%
Neuropathy	14%	15%	44%	45%	60%	62%
Amputation	0%	0%	3%	4%	9%	9%
Macular Edema Left	12%	14%	30%	31%	42%	41%
Macular Edema Right	12%	15%	30%	32%	41%	43%
Nonproliferative Retinopathy Left	17%	20%	44%	47%	48%	48%
Nonproliferative Retinopathy Right	17%	18%	46%	47%	46%	46%
Mortality Rate	1%	1%	1%	3%	2%	3%
Discounted Total Utility Score	0.60	0.47	4.99	4.56	8.35	7.94
Discounted Total Costs	$5,052	$7,017	$44,156	$71,860	$91,258	$115,494

The differences between the depressed and non-depressed patients become more pronounced in period 10, as a greater share of non-depressed patients remain in the healthy state

of the multiple disease sub-models compared to the depressed patients. Differences seem to be less pronounced in the retinopathy sub-model, whereas for the cardiovascular disease model, the changes in disease state rates is much more tangible. Again, the cost and utility differences move in the expected direction, with the total discounted utility score in period ten 0.43 points lower for depressed individuals and total discounted costs approximately 63% higher. In period 10, we see the proportion of the overall cohort that is depressed rise to 28%. Finally, in period 20, we see that there is a much more even gender split in the cohort when compared to period 1. The mortality rate gradually becomes larger for the population with depression when compared to the non-depressed population. This difference in mortality rates provides evidence that our modifications to the complication transition probabilities results in increased mortality for the population with depression. In period 10, the mortality rate was 1.86 times as high for the population with both diabetes and depression compared to the population with diabetes alone. Of the 2,947 deaths that occurred in our cohort across the 20 years, 73% of them were for individuals who had depression in at least 1 cycle.

Our simulation results suggest our modifications to the existing Michigan Model for Diabetes have resulted in the expected changes. Naturally, we are not 100% accurate in the assignment of depression status for every single individual in the simulation cohorts. Applying the prediction algorithm to the RAND HRS and NHANES datasets, we are able to characterize the individuals we correctly classify compared to those we misclassify. In both populations, the individuals we predict to be depression-free but who actually have depression are on average younger and healthier than the population we accurately classify as having depression. Specifically, in both the RAND HRS and NHANES samples, this population misidentified as being depression-free had lower BMI's on average, as well as much lower rates of heart disease,

stroke, and hypertension. So our prediction algorithm seems to primarily misclassify those individuals who have few comorbidities and complications, but still have depressive symptomology. If the individuals we identify with our model as having depression are sicker than the true population with depression, then our model could be biased by overestimating the benefits of treatment since we would avert more costs and complications by treating more severe patients than what would be the situation in reality. Conversely, we may also not capture the benefit of treating less severe patients, delaying their development of diabetic complications and comorbidities would add more years of healthy life. Therefore, the direction of this bias is not identifiable. But we are not interested in the exact accuracy at the individual level. Instead, the purpose of the modifications to the MMD are to aid population-level analyses. Our estimate of depression prevalence from our simulation is 17% in year 1, which is within the ranges described in the literature. We can be confident that these positive depression status indicators are being applied more to females, and individuals with hypertension, or histories of stroke or heart disease. Since all these variables have been strongly correlated with higher incidences of depression among individuals with diabetes in the literature, we are confident that some of our predicted individuals in each cohort who have depression are actually at a higher risk of depression.

Comparing results from the simulation using our modifications to the MMD with a parallel simulation with the original MMD, we find that the individuals who avoid cardiovascular disease and nephropathy have slightly lower BMI's, slightly lower systolic and diastolic blood pressures, and lower rates of smoking in our modified version. Similarly, in the updated MMD, the individuals that died over the 15 years of simulation were on average older (4 years) and had slightly higher BMI's than those that died over the 15 years of simulation in the

unmodified MMD. These comparisons suggest that our modifications may help better separate out individuals prone to complications and mortality than the original model, though the differences were generally quite small in magnitude.

Model Uncertainty

In order to incorporate our prediction model uncertainty, we use the results from bootstrapping our model 1,000 times to calculate the 95% confidence interval around the predicted probability threshold for assigning depression status. This process allows us to vary the cutoff point for the assignment of a positive depression status, where a lower threshold will result in more individuals as depressed, and a higher one will be a more stringent standard. Building this variability into our Michigan Model for Diabetes modifications allows future studies to determine the level of flexibility needed in depression status assignment. To calculate the bounds of the threshold point estimate, we calculate the predicted probabilities and then bootstrap the cut-point determination equation 1,000 times. Our resulting 95% confidence interval for this parameter was 0.4458 to 0.5569. Another source of uncertainty lies in our parameter estimate of the utility decrement associated with depression when using the QWB-SA scale. Accordingly, we run simulations for 1,000 individuals over 20 years where we vary the utility decrement for individuals with depression from -0.05 to -0.16. These simulations provide us with an idea of how the effect of depression can vary if there are environments that increase or decrease the utility decrement associated with depression when coexistent with diabetes. Again, our simulation results confirm our expectations, as an increase in the utility decrement (-0.16) was associated with lower average utility values for the population of depressed patients at all time periods (Figure 3.8).

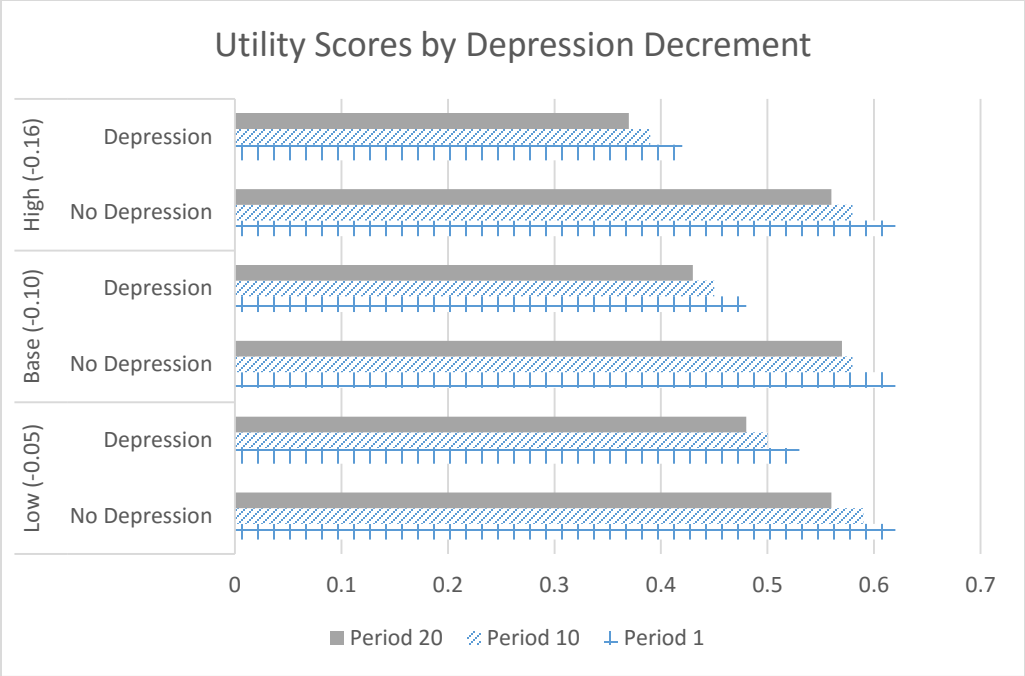


Figure 3.7 - Utility Sensitivity Analysis

Model Implementation Limitations

Our assumption of depression leading to increased complication risks directly operates outside of the individual-level risk factors that underlie the baseline transition probabilities of these complications and comorbidities. As mentioned before, the behaviors associated with depression can lead to poor diet, lower levels of physical activity, and poor self-care practices such as self-monitoring of blood glucose and medication adherence. Consequently, we could expect individuals with depression to experience changes in their blood pressure, lipid levels, smoking status, and body-mass index. Worse levels of all these risk factors could the increase the probability of developing any of the complications. Due to both data limitations as well as collinearity, we decide not to incorporate these mediating risk factor variables, and instead directly increase the transition probabilities into the complications and comorbidities. It is

unclear if the collinearity between some of these risk factors would lead to multiplicative effects of depression in the model if we were to model the increased risk of complications through these individual risk factors. Therefore, we believe that directly modifying the transition probabilities of these complications captures the effect we want, without the risk of inaccurately identifying changes to diabetes progression for individuals with depression. With future research, we may be able to identify the effect of depression on each of these risk factors independently, at which point we could then change the model structure of these modifications.

Another limitation of our implementation methodology is that it is not readily apparent which complication precedes the development of another one. Did a cardiovascular event lead to the depression, or did the depression lead to a cardiovascular event? This level of granularity is currently nonexistent with the way we have added depression to the Michigan Model for Diabetes. In future iterations, we may be better able to discern cause and effect relationships between the complications and depression status. The cyclical and interdependent nature of these complications and depressive symptoms would seem to prohibit the identification of truly independent effects of risk factors on depression development.

Similarly, depression among individuals with diabetes can have significant negative effects on individual medication adherence and treatment compliance [14]. Further research will need to be done to identify the approximate drop in treatment compliance associated with depression, as the MMD allows for variation in this parameter in simulations. This drop in compliance would increase the probability of transitioning into advanced disease states, but until there is more information available that would allow for the parameterization of this within the construct of the microsimulation model, we cannot account for this effect.

The data sources we use to develop the clinical prediction model allow us to construct depression status based on a threshold of responses to a survey of depressive symptomology. Accordingly, we are actually predicting the development of enough depressive symptomology that could lead to a depression diagnosis, instead of predicting actual depression. Since the outcome variable is a depression scale and not the actual clinical diagnosis of depression, there is a slight disconnect between what we want to model (actual depression) and what we actually predict (depression symptomology). The prediction model is still useful, as the responses to various depressive symptomology surveys are usually a primary input for the diagnosis of clinical depression. Another limitation is that in the studies of depression prevalence, as well as the Health and Retirement Study, the determinants of the depression outcome variable (scores on depressive symptomology survey), do not necessarily take into account whether or not an individual is currently receiving treatment for depression. Without a separation of those individuals who are receiving treatment for depression and still have depressive symptomology from those individuals who have not received any depression treatment but have depressive symptomology, we end up making predictions using this mixed population. Ideally, we would be able to predict depression status using data on individuals who are not receiving any treatment for depression but have a diagnosis of depression. Since we are unable to capture the individuals who have had depression but went through successful treatment with our dataset, this may bias our model by only identifying the more severe cases of depression and missing the benefit that can be accrued by treating patients who are less severe. Conversely, our model may overstate the benefits of treatment as we avert the more severe complications and costs instead of capturing the incremental differences of averting lower health care costs and less severe complications with treatment. Thus the direction of this bias is not readily apparent.

Furthermore, the dichotomization of depression status prevents us from analyzing the relationship between depression and pre-existing diabetes with more granularity. Previous work has documented the differences in the magnitude of relationships between predictor variables and major vs. minor depression among patients with diabetes [38]. Accordingly, there should be differences in the predictor variables as well as the transition probabilities for individuals with major vs. minor depression within the context of diabetes. Therefore, our estimates operate under the assumption of a population with major or minor depression when transition probabilities are adjusted. The direction of this bias is unclear since we cannot discern how the distribution of major vs. minor depression in our analytic samples compares to the distribution of the sample used to develop the estimates of increased complication risk by Lin et al. in their analysis [68]. As the current literature surrounding the increased risks of diabetes complications and comorbidities uses the dichotomous classification instead of a graded scale of depression severity, our implementation required that used a binary depression status variable. Future work with other sources of diabetes cohort data may be able to explore the importance of issues and improve the robustness of the model.

There is little existing evidence to predict which individuals in a cohort would respond to usual care treatment. Accordingly, we have to assume that all individuals have equal likelihood of being successfully treated. This assumption simplifies the effect of treatment, but is needed as there is not sufficient data to account for previous depression history. As mentioned earlier, we did not increase the independent risk of mortality from depression, but as more data on this relationship becomes available, we can modify as needed. With our current modifications, the population predicted to have depression experiences higher rates of mortality than the non-depressed population. This effect is mediated by the increased risk of complications for

depressed patients, who then subsequently face higher mortality risks. Another limitation of our MMD modifications is that our depressed population is disproportionately female. Although the prevalence of depression is significantly increased among females, our model seems to overemphasize this. Future research will work to calibrate the model to better match the gender split in patients with depression and diabetes.

Finally, our prediction model and the subsequent implementation into the microsimulation model operate uniformly. If an individual has depression for multiple years, the utility loss stays constant and the equation predicting depression status in subsequent cycles does not change. This simplification does not allow us to accurately model how there may be time-dependency determinants of both depression status as well as the associated utility loss of depression. These assumptions allow for simpler model structures, and may only be limiting in a few number of individuals in each cohort. Future work may explore a cycle-dependent utility loss function as well as a cycle-dependent change in depression prediction. Nevertheless, we believe our analysis takes an important step towards modeling depression among individuals with diabetes.

Conclusions

In this chapter we present results from the development of a depression prediction model that permits us to predict the depression status of every individual in a cohort of patients with diabetes. Based on their sex, body-mass index, hypertension status, history of stroke, history of heart disease, and duration of diabetes, we calculate a predicted probability of depression that was moderately discriminatory in identifying depression status based on data from the Health and Retirement Survey as well as the National Health and Nutrition Examination Survey. We then implement this prediction model into the Michigan Model for Diabetes, a microsimulation

model that allows users to study the progression of diabetes among a cohort of individuals. The Michigan Model for Diabetes models the development of neuropathy, nephropathy, retinopathy, cardiovascular disease, cerebrovascular disease, as well as direct medical costs and quality-of-life scores. The addition of a depression prediction algorithm and subsequent characterization of each cohort member as either depressed or depression-free in each cycle allowed for changes to the microsimulation model. Those with a positive depression status experienced increases in their transition probabilities into each of the modeled disease processes. This was based on the literature documenting the increased risk of both micro-vascular and macro-vascular complications for patients with both diabetes and depression when compared to patients with only diabetes. Concomitant with these increased transition probabilities was an associated decrease in the quality-of-life estimates from the simulation model whenever an individual had depression. The utility loss associated with depression is considered to be independent of the increases in complication and comorbidity development that diabetes patients with depression experience. We also build a usual care treatment parameter into the model to simulate the successful treatment of some of the individuals with depression every year so that depression is not an absorbing state. Finally, to build in parameter uncertainty with our modifications, we add ranges of cutoff thresholds to dichotomize the predicted probabilities of depression, as well as the range of utility decrements that could be associated with depression development.

The sensitivity of our prediction model was between 61-62%, the specificity was between 54-56%, and the area under the receiver operating characteristic curve was between 0.59-0.62. In a study predicting depression onset among the general population across multiple countries, King et al. found that sex, age, education, lifetime depression history, family history of psychological difficulties and scores on the Short Form 12 physical and mental health components were

significant predictor variables [85]. Their AUROC statistic ranged from 0.71-0.82, and the authors wanted to maximize specificity to reduce false positives, so they report sensitivity ranges from 32%-77%, and specificity ranges from 85-90%. Another study predicting onset of depression had a AUROC range of 0.70-0.80 when using data from electronic health records as predictor variables [86]. Both these studies had higher AUROC statistics, but they also were predicting the development of depression in the general population as opposed to a population with diabetes. Furthermore, many of the prediction models for depression incorporate previous depression history into their models. Although our model has worse discrimination than these models, we believe that within the context of pre-existing diabetes and no knowledge of previous depression experience, our model is beneficial in terms of identifying some of the patients that should have depression symptomology. Future work will investigate improvements of this prediction model to improve the clinical accuracy of the Michigan Model for Diabetes.

Our modifications to the Michigan Model for Diabetes resulted in simulation outcomes that matched expectations. Individuals with depression had higher complication rates, lower annual utility scores, and increased health care costs, three of the main distinguishing characteristics of this patient population when compared to patients with diabetes alone. The increases in the direct health care costs for individuals with depression when compared to those that were depression-free were in the ranges suggested by the literature. While there were limitations to our analyses, in both prediction model development as well as implementation in the microsimulation model, we think our analyses allow for future research to use a modeling approach to study the intersection of depression and diabetes.

References

1. *National Diabetes Statistics Report: Estimates of Diabetes and Its Burden in the United States, 2014.* , C.f.D.C.a. Prevention, Editor.: Atlanta, GA.

2. Herman WH, Rothberg AE, Prevalence of diabetes in the united states: A glimmer of hope? JAMA, 2015. 314(10): p. 1005-1007.

3. Menke A, et al., Prevalence of and trends in diabetes among adults in the united states, 1988-2012. JAMA, 2015. 314(10): p. 1021-1029.

4. American Diabetes Association, Economic Costs of Diabetes in the U.S. in 2012. Diabetes Care, 2013. **36**(4): p. 1033-1046.

5. Inzucchi SE, et al., Management of hyperglycemia in type 2 diabetes: a patient-centered approach. Diabetes Care, 2012. 35: 1364-1379.

6. *Major Depression Among Adults*. [cited 2016 March 1]; Available from: http://www.nimh.nih.gov/health/statistics/prevalence/major-depression-among-adults.shtml.

7. Greenberg PE, et al., The economic burden of adults with major depressive disorder in the United States (2005 and 2010). J Clin Psychiatry, 2015. 76(2): p. 155-62.

8. Talbot F, Nouwen A, A review of the relationship between depression and diabetes in adults: is there a link? Diabetes Care, 2000. 23(10): p. 1556-1562.

9. Anderson RJ, et al., The prevalence of comorbid depression in adults with diabetes: a meta-analysis. Diabetes Care, 2001. 24(6): p. 1069-1078.

10. Katon WJ, The comorbidity of diabetes mellitus and depression. Am J Med, 2008. 121(11 Suppl 2): p. S8-15.

11. Gavard JA, Lustman PJ, and Clouse RE, Prevalence of depression in adults with diabetes. An epidemiological evaluation. Diabetes Care, 1993. 16(8): p. 1167-78.

12. Fisher L, et al., Contributors to depression in latino and european-american patients with type 2 diabetes. Diabetes Care, 2001. 24(10): p. 1751-1757.

13. Li C, et al., Prevalence of depression among u.s. adults with diabetes. findings from the 2006 behavioral risk factor surveillance system. Diabetes Care, 2008. 31(1): p. 105-107.

14. DiMatteo MR, Lepper HS, Croghan TW, Depression is a risk factor for noncompliance with medical treatment: meta-analysis of the effects of anxiety and depression on patient adherence. Arch Intern Med, 2000. 160(14): p. 2101-7.

15. Katon WJ, et al., Cardiac risk factors in patients with diabetes mellitus and major depression. J Gen Intern Med, 2004. 19(12): p. 1192-9.

16. de Groot M, et al., Association of depression and diabetes complications: a meta-analysis. Psychosom Med, 2001. 63(4): p. 619-30.

17. Huang Y, et al., Collaborative care for patients with depression and diabetes mellitus: a systematic review and meta-analysis. BMC Psychiatry, 2013. 13(1): p. 1-11.

18. Jeeva F, et al., Is treatment of depression cost-effective in people with diabetes? A systematic review of the economic evidence. Int J Technol Assess Health Care, 2013. 29(4): p. 384-91.

19. Gross JL, et al., Diabetic nephropathy: diagnosis, prevention, and treatment. Diabetes Care, 2005. 28(1): p. 164-176.

20. Dokken BB, The pathophysiology of cardiovascular disease and diabetes: beyond blood pressure and lipids. Diabetes Spectrum, 2008. 21(3): p. 160-165.

21. Brigitta B, Pathophysiology of depression and mechanisms of treatment. Dialogues Clin Neurosci, 2002. 4(1): p. 7-20.

22. Hasler G, Pathophysiology of depression: do we have any solid evidence. World Psychiatry, 2010. 9(3): p. 155-61.

23. Hirschfeld RMA, The comorbidity of major depression and anxiety disorders: recognition and management in primary care. Prim Care Companion J Clin Psychiatry, 2001. 3(6): p. 244-54.

24. Licht-Strunk E, et al., Outcome of depression in later life in primary care: longitudinal cohort study with three years' follow-up. BMJ, 2009. 338.

25. Simon GE, VonKorff M, Barlow W, Health care costs of primary care patients with recognized depression. Arch Gen Psychiatry, 1995. 52(10): p. 850-6.

26. Kessler RC, et al., The epidemiology of major depressive disorder: Results from the national comorbidity survey replication (ncs-r). JAMA, 2003. 289(23): p. 3095-3105.

27. Rutledge T, et al., Depression in heart failure a meta-analytic review of prevalence, intervention effects, and associations with clinical outcomes. J Am Coll Cardiol, 2006. 48(8): p. 1527-37.

28. Mezuk B, Eaton WW, Golden SH, Depression and osteoporosis: epidemiology and potential mediating pathways. Osteoporos Int, 2008. 19(1): p. 1-12.

29. Fan AZ, et al., Depression and anxiety associated with cardiovascular disease among persons aged 45 years and older in 38 states of the United States, 2006. Prev Med, 2008. 46(5): p. 445-50.

30. Lustman PJ and Clouse RE, Treatment of depression in diabetes: impact on mood and medical outcome. J Psychosom Res, 2002. 53(4): p. 917-24.

31. Renn BN, Feliciano L, Segal DL, The bidirectional relationship of depression and diabetes: a systematic review. Clin Psychol Rev, 2011. 31(8): p. 1239-46.

32. Kinder LS, et al., Depressive symptomatology and coronary heart disease in Type I diabetes mellitus: a study of possible mechanisms. Health Psychol, 2002. 21(6): p. 542-52.

33. Knol MJ, et al., Depressive symptoms in subjects with diagnosed and undiagnosed type 2 diabetes. Psychosom Med, 2007. 69(4): p. 300-5.

34. Knol MJ, et al., Depression as a risk factor for the onset of type 2 diabetes mellitus. A meta-analysis. Diabetologia, 2006. 49(5): p. 837-45.

35. Mezuk B, et al., Depression and type 2 diabetes over the lifespan: a meta-analysis. Diabetes Care, 2008. 31(12): p. 2383-90.

36. Ciechanowski PS, Katon WJ, Russo JE, Depression and diabetes: impact of depressive symptoms on adherence, function, and costs. Arch Intern Med, 2000. 160(21): p. 3278-85.

37. Barbour KA, Blumenthal JA, Exercise training and depression in older adults. Neurobiol Aging, 2005. 26 Suppl 1: p. 119-23.

38. Katon WJ, et al., Behavioral and clinical factors associated with depression among individuals with diabetes. Diabetes Care, 2004. 27(4): p. 914-20.

39. Black SA, Markides KS, Ray LA, Depression predicts increased incidence of adverse health outcomes in older mexican americans with type 2 diabetes. Diabetes Care, 2003. 26(10): p. 2822-2828.

40. Goldney RD, Phillips PJ, Fisher LJ, Wilson DH. Diabetes, depression, and quality of life. Diabetes Care, 2004. 27(5): 1066-1070.

41. Golden SH, et al., Examining a bidirectional association between depressive symptoms and diabetes. JAMA, 2008. 299(23): p. 2751-9.

42. Bjorntorp P, Do stress reactions cause abdominal obesity and comorbidities? Obes Rev, 2001. 2(2): p. 73-86.

43. Palmer AJ, et al., Computer modeling of diabetes and its complications: a report on the Fifth Mount Hood challenge meeting. Value Health, 2013. 16(4): p. 670-85.
44. Ye W, et al., The michigan model for coronary heart disease in type 2 diabetes: development and validation. Diabetes Technol Ther, 2015. 17(10): p. 701-11.

45. Herman WH, et al., Early detection and treatment of type 2 diabetes reduce cardiovascular morbidity and mortality: a simulation of the results of the anglo-danish-dutch study of intensive treatment in people with screen-detected diabetes in primary care (addition-europe). Diabetes Care, 2015.

46. Zhou H, et al., A computer simulation model of diabetes progression, quality of life, and cost. Diabetes Care, 2005. 28(12): p. 2856-63.

47. Isaman DJ, Barhak J, Ye W, Indirect estimation of a discrete-state discrete-time model using secondary data analysis of regression data. Stat Med, 2009. 28(16): p. 2095-115.

48. Barhak J, et al., Chronic disease modeling and simulation software. J Biomed Inform, 2010. 43(5): p. 791-9.

49. Wolfe F, Michaud K, Predicting depression in rheumatoid arthritis: The signal importance of pain extent and fatigue, and comorbidity. Arthritis Care & Research, 2009. 61(5): p. 667-673.

50. Nordin K, et al., Predicting anxiety and depression among cancer patients: a clinical model. European Journal of Cancer, 2001. 37(3): p. 376-384.

51. Van Voorhees BW, et al., Predicting future risk of depressive episode in adolescents: the chicago adolescent depression risk assessment (cadra). The Annals of Family Medicine, 2008. 6(6): p. 503-511.

52. Kaur G, et al, Depression, anxiety and stress symptoms among diabetics in Malaysia: a cross sectional study in an urban primary care setting. BMC Fam Pract, 2013. 14: p. 69.

53. Jin H, Wu S, and Di Capua P, Development of a clinical forecasting model to predict comorbid depression among diabetes patients and an application in depression screening policy making. Preventing Chronic Disease, 2015. 12: p. E142.

54. Burcusa SL, Iacono WG, Risk for recurrence in depression. Clin Psychol Rev, 2007. 27(8): p. 959-85.

55. Justice AC, Covinsky KE, Berlin JA, Assessing the generalizability of prognostic information. Ann Intern Med, 1999. 130(6): p. 515-24.

56. Steyerberg EW, et al., Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology, 2010. 21(1): p. 128-38.

57. Heisler M, et al., Does physician communication influence older patients' diabetes self-management and glycemic control? results from the health and retirement study (hrs). The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 2007. 62(12): p. 1435-1442.

58. Nothwehr F, Stump T, Health-promoting behaviors among adults with type 2 diabetes: findings from the health and retirement study. Preventive Medicine, 2000. 30(5): p. 407-414.

59. Heisler M., et al., Mechanisms for racial and ethnic disparities in glycemic control in middle-aged and older americans in the health and retirement study. Archives of Internal Medicine, 2007. 167(17): p. 1853-1860.

60. Blaum CS, et al., Clinical complexity in middle-aged and older adults with diabetes: the health and retirement study. Medical care, 2010. 48(4): p. 327-334.

61. Steffick DE, Documentation of affective functioning measures in the health and retirement study. 2000, University of Michigan: Survey Research Center.

62. Zuithoff NP, et al., A clinical prediction rule for detecting major depressive disorder in primary care: the PREDICT-NL study. Fam Pract, 2009. 26(4): p. 241-50.

63. Lorant V, et al., Socioeconomic Inequalities in Depression: A Meta-Analysis. American Journal of Epidemiology, 2003. 157(2): p. 98-112.

64. Steyerberg EW, Vergouwe Y, Towards better clinical prediction models: seven steps for development and an ABCD for validation. Eur Heart J, 2014. 35(29): p. 1925-31.

65. Roy T, Lloyd CE, Epidemiology of depression and diabetes: a systematic review. J Affect Disord, 2012. 142 Suppl: p. S8-21.

66. Almeida OP, et al., Duration of diabetes and its association with depression in later life: The Health In Men Study (HIMS). Maturitas. 86: p. 3-9.

67. Siu, AL and U.S.P.S.T.F., Screening for depression in adults: Us preventive services task force recommendation statement. JAMA, 2016. 315(4): p. 380-387.

68. Lin EH, et al., Depression and advanced complications of diabetes: a prospective cohort study. Diabetes Care, 2010. 33(2): p. 264-9.

69. Schulz R, et al., Association between depression and mortality in older adults: The cardiovascular health study. Archives of Internal Medicine, 2000. 160(12): p. 1761-1768.

70. Katon WJ, et al., The association of comorbid depression with mortality in patients with type 2 diabetes. Diabetes Care, 2005. 28(11): p. 2668-2672.

71. Bruce DG, et al., A prospective study of depression and mortality in patients with type 2 diabetes: the Fremantle Diabetes Study. Diabetologia, 2005. 48(12): p. 2532-2539.

72. Sundaram M, et al., Quality of life, health status and clinical outcomes in Type 2 diabetes patients. Qual Life Res, 2007. 16(2): p. 165-77.

73. Caruso LB, et al., What can we do to improve physical function in older persons with type 2 diabetes? J Gerontol A Biol Sci Med Sci, 2000. 55(7): p. M372-7.

74. Ali S, et al., The association between depression and health-related quality of life in people with type 2 diabetes: a systematic literature review. Diabetes/Metabolism Research and Reviews, 2010. 26(2): p. 75-89.

75. Pyn JM, et al., Use of the quality of well-being self-administered version (QWB-SA) in assessing health-related quality of life in depressed patients. Journal of Affective Disorders, 2003. 76(1–3): p. 237-247.

76. Fryback DG, et al., The beaver dam health outcomes study: initial catalog of health-state quality factors. Medical Decision Making, 1993. 13(2): p. 89-102.

77. Pyne JM, et al., Assessment of the quality of life of patients with major depression. Psychiatr Serv, 1997. 48(2): p. 224-30.

78. Kaplan RM, et al., The Quality of Well-Being Scale: critical similarities and differences with SF-36. Int J Qual Health Care, 1998. 10(6): p. 509-20.

79. Coffey JT, et al., Valuing health-related quality of life in diabetes. Diabetes Care, 2002. 25(12): p. 2238-2243.

80. Zhang P, et al., Health utility scores for people with type 2 diabetes in U.S. managed care health plans: results from Translating Research Into Action for Diabetes (TRIAD). Diabetes Care, 2012. 35(11): p. 2250-6.

81. Johnson JA, et al., Collaborative care versus screening and follow-up for patients with diabetes and depressive symptoms: results of a primary care-based comparative effectiveness trial. Diabetes Care, 2014. 37(12): p. 3220-6.

82. Bogner HR., et al., Integrated management of type 2 diabetes mellitus and depression treatment to improve medication adherence: a randomized controlled trial. Ann Fam Med, 2012. 10(1): p. 15-22.

83. Unutzer J et al., Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. Jama, 2002. 288(22): p. 2836-45.

84. Katon W, et al., Cost-effectiveness and net benefit of enhanced treatment of depression for older adults with diabetes and depression. Diabetes Care, 2006. 29(2): p. 265-70.

85. King M, et al., Development and validation of an international risk prediction algorithm for episodes of major depression in general practice attendees. Arch Gen Psychiatry, 2008. 65(12): 1368-1376.

86. Huang SH, et al., Toward personalizing treatment for depression: predicting diagnosis and severity. J Am Med Inform Assocm 2014. 21: 1069-1075.

Chapter 3 Appendix

Here we show the variable narrowing process used in our model development. We first started with the comprehensive list of variables that are associated with depression among individuals with diabetes. This was taken from the Fisher et al [12].

Table 1—*An overview of factors affecting the prevalence of depression in patients with diabetes*

Patient demographics
- Sex: generally higher rates for females (4, 7–12)
- Race/ethnicity: higher rates for minorities (8, 13–15)
- Marital status: higher rates for singles (4, 16)
- Age: mixed findings, somewhat higher rates in the middle years and somewhat lower rates in the older years (4, 11, 17, 18, 40)
- Social class (education, income): higher rates at lower levels of social class (8, 12, 19, 20, 57)
- Employment status: higher rates for the unemployed (12)

Disease characteristics
- Number of comorbidities/complications: mixed findings with a trend toward higher rates with more comorbidities/complications (4, 11, 12, 18, 20–32, 40, 57)
- Type 1 versus type 2 diabetes: mixed findings, may be related to presence or absence of lifetime history of depression (33, 57)
- Degree of disability: high rates with high disability (24)
- Pain, discomfort, physical impairment: high rates associated with each (42)

Social support: low depression associated with high social support (17)

General psychological disturbance: depression associated with.
- Anxiety (57)
- Health perceptions (34)
- Negative life events (35)
- Adverse social circumstances (12)
- Illness intrusiveness (36)
- Affect and quality of life (16, 17, 37, 38)
- Well-being (39)

Then, we had to see what variables were available in our primary dataset:

RAND HRS Dataset:

1) Demographic Information – gender, age, race, marital status, education level, income, employment status, survey year
2) Health Care Information: medical utilization and expenditures, self-rated health, BMI, along with condition-specific health history. History or current development of hypertension, diabetes, cancer, lung disease, heart disease, stroke, psychiatric illnesses, arthritis, back pain, and ulcers.

And finally, we then had to account for the pertinent variables that were tracked in each cycle of the Michigan Model for Diabetes

Michigan Model for Diabetes:

1) Demographic Information – gender, age, 2-category race, BMI, duration of diabetes
2) Disease Status – cholesterol and blood pressure levels, HbA1c levels, disease states within cerebrovascular disease, coronary heart disease, retinopathy, nephropathy and neuropathy

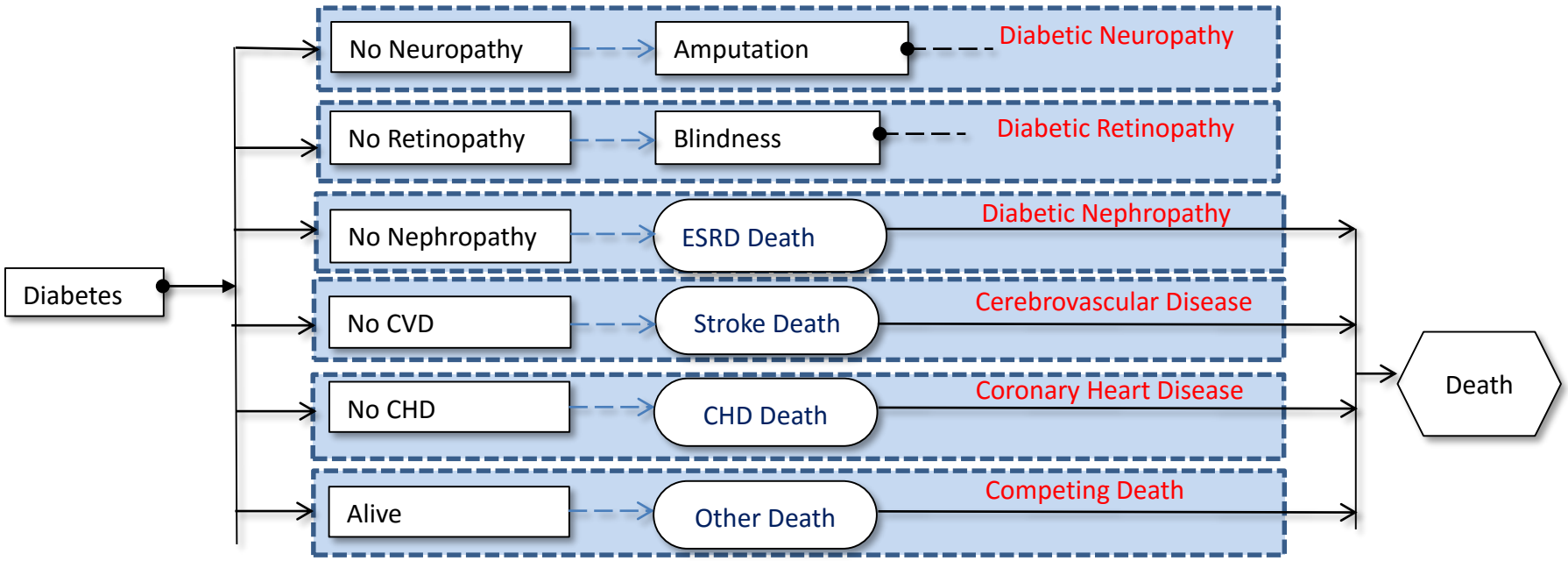


Figure A1. Overall Structure of Michigan Model for Diabetes.

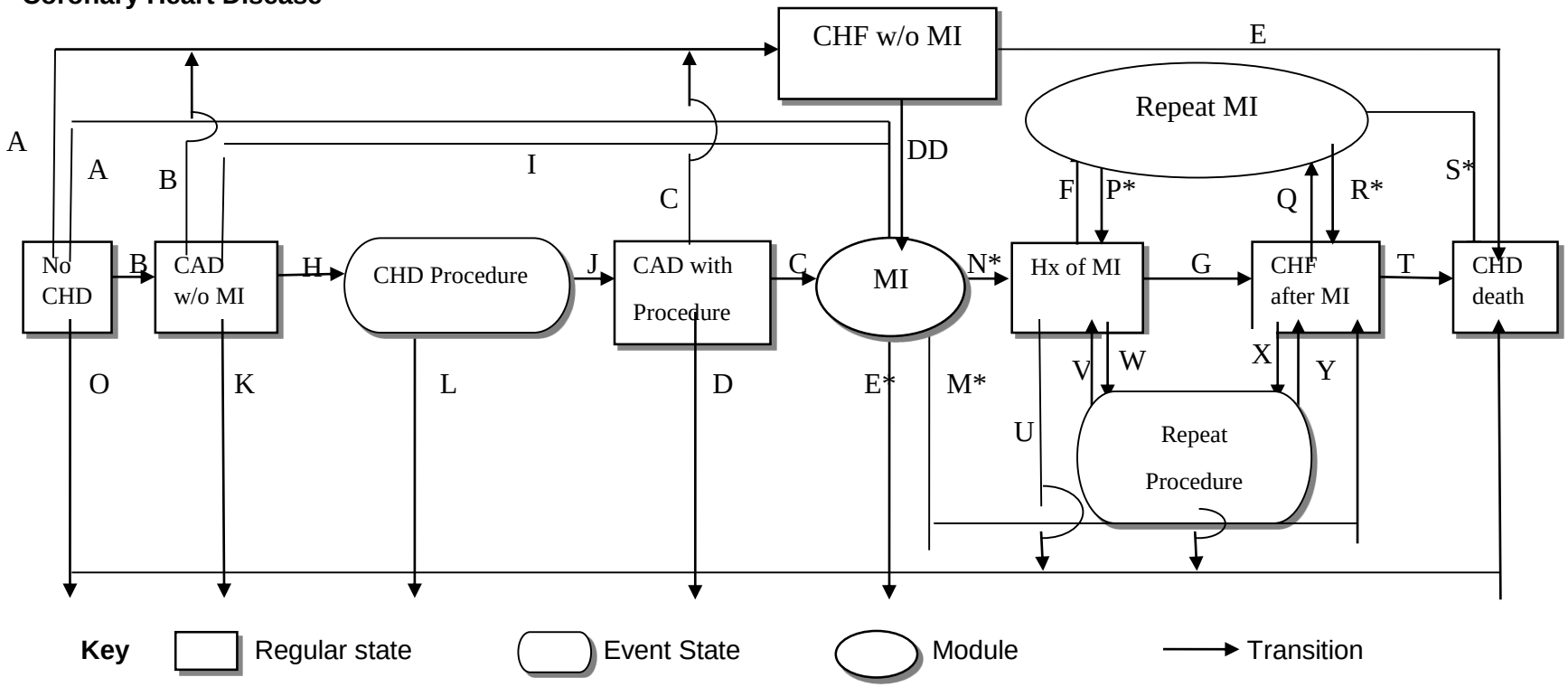
Key Regular state Event State Module Terminal State Transition

Hidden transitions shown in Figures 2 to 8 Splitting transition to multiple sub-processes

No transition initiated from the initial state, but the sub-process can be ended due to another sub-process reaches the terminal state

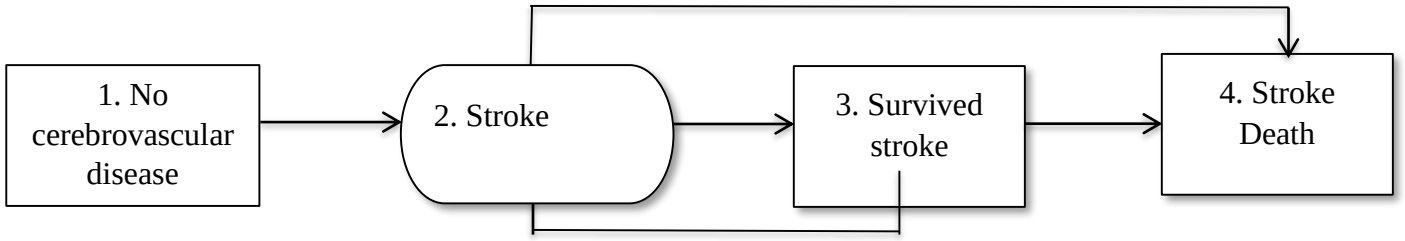
Nested parallel sub-processes

Coronary Heart Disease

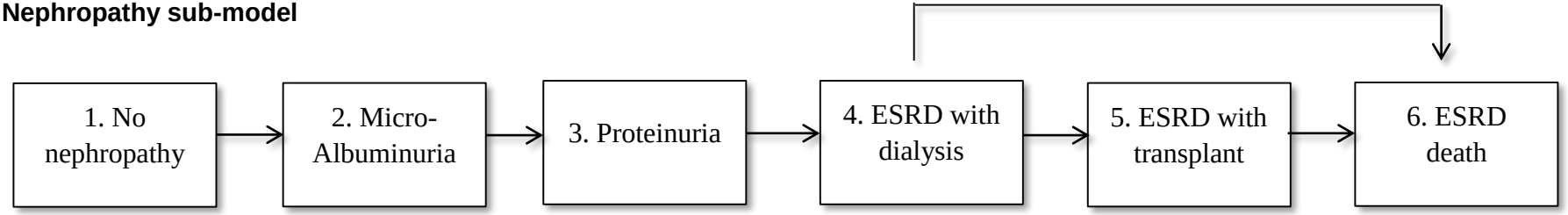


Coronary heart disease states and progression. CHD=coronary heart disease, CAD=coronary artery disease, CHF w/o MI= congestive heart failure without MI, MI=myocardial Infarction, CHF=congestive heart failure after experience of MI, Hx=history, w/o=without, CHD Procedure=revascularization procedure.

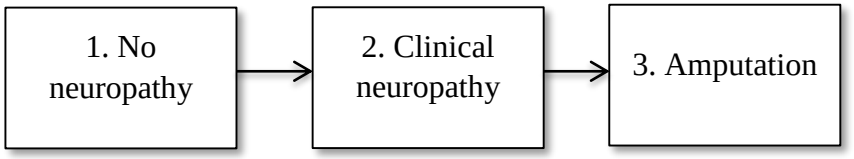
Cerebrovascular disease



Nephropathy sub-model

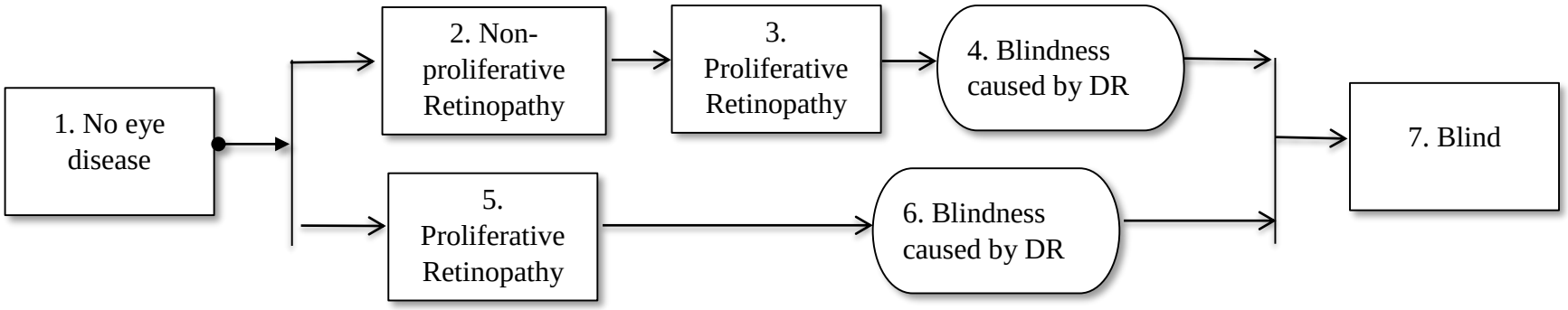


Neuropathy



Retinopathy

Two eyes are modeled separately and assume to be independent. Retinopathy, macular edema are two parallel sub-sub-processes.



Transition Probability Changes

No Nephropathy to Microalbuminuria:

Iif(Depression, (Gaussian(1.36,.1775)*(0.0509/(0.8 + (Gaussian(1.36,.1775)*0.2)))), (0.0509/(0.8 + (Gaussian(1.36,.1775)*0.2))))

No Neuropathy to Clinical Neuropathy

Iif(Depression, (Gaussian(1.36, .1775)*(0.0518/(0.8 + (Gaussian(1.36, .1775)*0.2)))), (0.0518/(0.8 + (Gaussian(1.36, .1775)*0.2))))

No Macular Edema Left to Macular Edema Left

Iif(Depression, (Gaussian(1.36, .1775)*(0.0308/(0.8 + (Gaussian(1.36, .1775)*0.2)))), (0.0308/(0.8 + (Gaussian(1.36, .1775)*0.2))))

No Proliferative Retinopathy to Nonproliferative Retinopathy Left

Iif(Depression, (Gaussian(1.36, .1775)*(Iif(Insulin,0.1140,0.0653))/(0.8 + (Gaussian(1.36, .1775)*0.2))),
(Iif(Insulin,0.1140,0.0653)/(0.8 + (Gaussian(1.36, .1775)*0.2))))

No Macular Edema Right to Macular Edema Right

Iif(Depression, (Gaussian(1.36, .1775)*(0.0308/(0.8 + (Gaussian(1.36, .1775)*0.2)))), (0.0308/(0.8 + (Gaussian(1.36, .1775)*0.2))))

No proliferative retinopathy to Nonproliferative Retinopathy Right

Iif(Depression, (Gaussian(1.36, .1775)*(Iif(Or(Insulin,BasalInsulin),0.1140,0.0653))/(0.8 + (Gaussian(1.36, .1775)*0.2))),
(Iif(Or(Insulin,BasalInsulin),0.1140,0.0653))/(0.8 + (Gaussian(1.36, .1775)*0.2)))

No CVD to Angina

Iif(Depression, (Gaussian(1.25, .135)*(Iif(Ge(Prob_From_No_CVD,1), Prob_From_No_CVD_To_Angina, (1-Exp(-3.78*OutcomeIHDcumHaz))*0.786)/(0.8 + (Gaussian(1.25, .135)*0.2)))), (Iif(Ge(Prob_From_No_CVD,1), Prob_From_No_CVD_To_Angina, (1-Exp(-3.78*OutcomeIHDcumHaz))*0.786)/(0.8 + (Gaussian(1.25, .135)*0.2))))

No CVD to CHFwoMI

Iif(Depression, (Gaussian(1.25,0.135)*(Iif(Ge(Prob_From_No_CVD,1), Prob_From_No_CVD_To_CHF, 1-Exp(-CHFHumHaz_CHSstudy2))/(0.8 + (Gaussian(1.25,0.135)*0.2)))), (Iif(Ge(Prob_From_No_CVD,1), Prob_From_No_CVD_To_CHF, 1-Exp(-CHFHumHaz_CHSstudy2))/(0.8 + (Gaussian(1.25,0.135)*0.2))))

No CVD to MI

Iif(Depression, (Gaussian(1.25,0.135)*(Iif(Ge(Prob_From_No_CVD,1), Prob_From_No_CVD_To_MI, (1-Exp(-OutcomeMIcumHazNoIHDNoHF*3.78))*0.183)/(0.8 + (Gaussian(1.25,0.135)*0.2)))), (Iif(Ge(Prob_From_No_CVD,1), Prob_From_No_CVD_To_MI, (1-Exp(-OutcomeMIcumHazNoIHDNoHF*3.78))*0.183)/(0.8 + (Gaussian(1.25,0.135)*0.2))))

No CVD to CVD Death

Iif(Depression, (Gaussian(1.25,0.135)*(Iif(Ge(Prob_From_No_CVD,1), Prob_From_No_CVD_To_Death, (1-Exp(-OutcomeMIcumHazNoIHDNoHF*3.78))*0.024)/(0.8 + (Gaussian(1.25,0.135)*0.2)))), (Iif(Ge(Prob_From_No_CVD,1), Prob_From_No_CVD_To_Death, (1-Exp(-OutcomeMIcumHazNoIHDNoHF*3.78))*0.024)/(0.8 + (Gaussian(1.25,0.135)*0.2))))

No CVD to Stroke

Iif(Depression, (1-Exp(-((Gaussian(1.25,.135))/(0.8 + (Gaussian(1.25,.135)*0.2)))*cumHazstroke*CVDDrugEffect)), (1-Exp(-(1/(0.8 + (Gaussian(1.25,.135)*0.2)))*cumHazstroke*CVDDrugEffect)))

Chapter 4
Using Modeling to Study the Cost-Effectiveness of the Collaborative Care Intervention Among Patients with Diabetes and Depression

Introduction

As the burden of chronic illness in the United States increases, the need to identify efficient treatment options will increase in importance [1]. Estimates suggest that the worldwide prevalence of diabetes will increase to 7.7% by the year 2030 and this growing disease burden is a clear example of the need for better treatment approaches to improve health and reduce costs [2]. One of the main factors underlying the necessity for better treatment among patients with diabetes is the increased risk of developing many associated complications and comorbidities. For example, studies have shown that the risk of depression among individuals with diabetes is twice as high as the general population [3]. Within this subpopulation of individuals with both diabetes and depression, a leading cause for concern is the greater likelihood of developing the other complications that accompany diabetes [4]. Patients with both diabetes and depression are characterized as having lower treatment adherence, worse overall disease management, higher complication rates, decreased quality-of-life, and increased health care costs. Accordingly, this patient population could benefit greatly from expanded efforts to treat their illnesses.

Prior research suggests that a collaborative care approach which integrates care by nurses and physicians can be effective in improving depressive symptoms among patients with diabetes [5, 6]. Implementation of this treatment approach has shown beneficial improvements in the short-term (9-24 months), with decreases in depressive symptoms compared to standard care

from physicians [6]. Unfortunately, the evidence thus far is limited by both follow-up length as well as sample generalizability. This evidence gap could benefit from modeling studies to improve our understanding of the potential costs and benefits of treating depression among patients with diabetes. In our analysis, we use a microsimulation model of diabetes that has been modified to predict and track depression status over time to study the treatment of depression among patients with diabetes. By using a modeling approach to study the collaborative care intervention approach for this patient population, we can generate estimates of the longer-term cost-effectiveness of this treatment.

Thus far, the literature surrounding the development, prevalence and treatment of depression among individuals with diabetes has been important in developing an understanding of this patient population. A variety of studies have researched the mechanisms underlying the relationship between depression and diabetes. Most of the evidence suggests a bidirectional mechanistic relationship between diabetes and depression, with both biological and psychosocial mediating factors. Reviews of the literature suggest that diabetes doubles the odds of depression, with increased prevalence among women [7]. With this increased prevalence of depression, the health care burden is also heightened. Patients with both diabetes and depression suffer from lower treatment adherence, worse lifestyle habits, increased probability and severity of diabetes symptoms and complications, and greater mortality [8, 9]. There is also evidence that the co-occurrence of depression and diabetes can have a synergistic effect on risk of poorer health outcomes among patients [8]. This synergistic effect may be explained in part by the decrease in treatment adherence and worse lifestyle characteristic of patients with both depression and diabetes.

Patients with both depression and diabetes incur greater health care costs and suffer from lower quality-of-life. Surveying primary care patients in the Group Health Cooperative, Ciechanowski et al. find that total, ambulatory, as well as primary care adjusted 6-month health care costs were significantly increased for high severity depression patients compared to those with low severity [10]. Specialty care costs were not increased across the depression spectrum. The authors conclude that since depressive symptom severity is associated with greater costs and lower adherence to treatment protocols, much of the detrimental effects of these two illnesses could be decreased through effective treatment of the depression [10]. In another study using the Group Health Cooperative population, Simon et al. found that costs for diabetes treatment, other medical costs, and total health care costs were approximately 70% greater in patients with depression. Total six-month health care costs were $2,241 greater in those with major depression and diabetes, compared to patients with only diabetes. This difference ranged from $500 for those with no other complications, to $3,000 for those with multiple complications. The authors find that those with three or more complications of diabetes accounted for 39% of costs while only being 18% of the sample. Mental health care costs were not a large portion of the increased total health care costs in this population; overall, their results documented a need for an increased focus on this population to improve disease management [11].

In 2012, Molosankwe et al. completed a systematic review of the economics behind the association between diabetes and depression [12]. They found 62 studies that met their inclusion criteria, and summarized resource utilization, health care costs, as well as the cost-effectiveness of interventions within this patient population. Their review of the utilization studies showed that much of the increased costs arose from general health care services, and not necessarily mental health services costs. They also documented the lower levels of treatment adherence and greater

risk of complications for patients with both diabetes and depression [12]. With these higher risks of other diabetic complications, studies have also shown parallel increases in the costs of care associated with such patients. The authors conclude that there is a need for modeling studies to evaluate the potential costs and benefits of interventions in a population of individuals with both diabetes and depression, as there seem to be no such studies in the United States [12].

Overall, it is evident that patients with diabetes and depression have increased health care costs and health care utilization. Although much of the evidence up to now has been compiled by the same research group with the same general patient population, this literature establishes the need for focused treatment interventions within this patient population to work towards better disease management. Considering that the investment for such disease management programs would most likely have to be financed by health plans, there would need to be a strong economic incentive for organizations to do so. Although there is a possibility that organizations may not be able to capture future returns on their investment if patients leave insurance plans, the case for disease management can be strong because of the economic, health, and quality-of-life benefits accrued [13].

There is considerable evidence and support for using a collaborative care approach to treat patients with both diabetes and depression. This approach entails physicians, a case manager, and a mental health specialist working closely with a patient to help improve their depressive symptomology [14]. In most of the studies done thus far, the intervention has been a one-time 12-month intervention that uses nurses and physicians to continuously interface with patients while using both pharmacotherapy and self-care therapy to improve patient health [6]. Many of these studies have documented beneficial outcomes, with improvements in depressive symptoms documented at the 12 to 24-month range. Evidence from studies of using collaborative

care among patients with depression (without diabetes) provide evidence of longer-term benefit up to 5 years [15]. Among patients with diabetes, studies suggest the collaborative care approach can lead to better depression outcomes.

Four studies have done an economic analysis of this type of intervention. Evaluating the cost-effectiveness of a 12-month stepped-care depression treatment program administered by nurse case managers for patients with diabetes and depressive symptoms in the primary care setting, Simon et al. found beneficial health and economic benefits [16]. Tracking Hopkins-Symptoms Checklist 20 Depression Scale (HSCL-20) depression scores over time along with depression-free days, as well as diabetes clinical outcomes, they found the intervention lowered depression scores, increased the number of depression-free days, and had no significant impact on HbA1c levels at both 12 and 24 months. While the intervention increased total health care costs in year 1, in the 2nd year, the intervention group had approximately $1,400 less in total health care spending. They calculated incremental cost-effectiveness ratios based on the number of depression free days, because they did not use any utility metric to determine overall effect on quality-adjusted life years (QALYs). The intervention was dominant, with the 95% confidence interval crossing the cost-axis, indicating a possibility of the intervention increasing costs while improving effectiveness. Although there is no set willingness-to-pay (WTP) for a depression-free-day, they show that their results are fairly robust across a range of economic values for this [16].

The Improving-Mood-Promoting Access to Collaborative (IMPACT) trial was a well-publicized study documenting the beneficial health and economic effects of a collaborative care depression intervention in elderly depressed patients [17]. A sub-group analysis of the effects of IMPACT on elderly patients with both diabetes and depression found it to be a high-value

investment. Similar to the previously mentioned study, HSCL-20 scores were determined throughout the intervention to determine the number of depression-free days during follow-up. Over the 24 months, the intervention group saved $896 compared to the control group, driven in a large part by lower inpatient costs. The authors mapped the number of depression-free days onto QALY gains based on previous metrics, determining that the incremental cost-effectiveness ratio (ICER) was between $198-$397 [18].

A study in 2012 evaluated the cost-effectiveness of a collaborative care intervention for individuals with depression and poorly controlled diabetes and/or coronary heart disease. Nurse case managers worked with primary care providers to administer treatment for multiple disease factors over 12 months, while outpatient, inpatient, and intervention costs were collected for all patients [19]. Again, the authors used depression-free days as well as any changes in HbA1c, systolic blood pressure, and LDL-C levels to estimate QALY gains. They found the intervention to be dominant compared to usual care when calculating the incremental outpatient cost per depression-free day, though the 95% confidence interval crossed the cost-axis. When they use their estimated QALY gains, they also find the intervention dominated the usual care strategy [19].

Another study evaluated the cost-effectiveness of a collaborative-care depression management intervention among a population of low-income, predominantly Hispanics with diabetes [20]. The intervention study provided evidence that evidence-based pharmacotherapy and/or psychotherapy had greater depression improvement than usual care, though there were no beneficial effects on glycemic control. Analyzing the cost-effectiveness of the intervention at the 18-month time period, they find the intervention ICER to be $4,053 compared with usual care.

They only included intervention costs in their calculations due to a lack of significant differences between treatment and control health care costs [20].

While these studies evaluating the cost-effectiveness of interventions to address the needs of patients with diabetes and depression are very informative, they suffer from a variety of limitations. First, few of these studies report any type of sensitivity analysis. It is difficult to gauge external validity without any indication of what variables could be driving their results. Even though most of the studies provide intervention costs, other parameters could be drastically different based on provider location or patient characteristics. Especially since many of these interventions occur on distinct patient populations that may or may not be representative of the general population, understanding how different variables could be driving the results is necessary evidence. Furthermore, none of these studies evaluate the long-term ICERs of any of these interventions, limiting their findings only to the follow-up time period of the study. It is foreseeable that there could be diminishing effects of the intervention over time, thus any health benefits may be reduced in the long-term. It is extremely important to study the long-term effects of these types of interventions, because if they do remain effective in the long-term, then there could be an even stronger case for implementation of similar interventions by health plans. Limiting analyses to short time periods prevents us from gauging the unbiased benefits of the intervention. Overall, there have been very few cost-effectiveness analyses of depression interventions among a patient population with diabetes.

A recently published systematic review evaluated the evidence on whether or not the treatment of depression in people with diabetes was cost-effective. The authors identified the four studies discussed here as the only true economic evaluations, and concluded that while these studies have shown positive results, there was a need for economic models to extrapolate the

results to more generalizable settings as well as longer time-horizons [21]. There is a clear need for modeling studies to fill the gaps in the current state of knowledge regarding such an intervention.

We evaluate the cost-effectiveness of implementing a 12-month stepped care intervention among patients with both diabetes and depression by analyzing health care costs and utility scores at multiple time points after intervention implementation. Initial intervention efficacy is based on existing trial data. The intervention lasts for 12-months, while outcome data exist for 12-60 months [15]. Therefore, we use a combination of the data from these studies as well as a variety of different waning functions to predict intervention effectiveness in the longer-term. Since there is little data supporting the long-term effectiveness of the collaborative care intervention, we perform sensitivity analyses on this parameter to estimate the variation in the cost-effectiveness. Comparing the discounted total direct medical costs and total health-related quality-of-life scores of cohorts simulated through either the collaborative care treatment or usual care comparator scenario, we generate estimates of the short and long-term cost-effectiveness of using a multidisciplinary approach to address the health care needs of patients with both diabetes and depression.

Methods

Overview of Approach

In this analysis, we use an existing microsimulation model of diabetes progression to study the cost-effectiveness of applying the collaborative care approach to treat depression among patients with diabetes. In order to build on the existing evidence, we predict the waning of treatment effectiveness up to 15 years and apply cost and utility estimates from the literature to calculate the incremental cost-effectiveness ratio of the collaborative care intervention when

compared to the standard usual care of depression among patients with diabetes. We use a health care payer perspective for our analysis, as that is the most relevant lens for the administration of an intervention such as the collaborative care approach.

Model Overview

We use the modified version of the Michigan Model for Diabetes (MMD) to evaluate the cost-effectiveness of treating depression among individuals with diabetes. The Michigan Model for Diabetes is a publicly available microsimulation discrete-time model that simulates the progression of diabetes by modeling the development of cardiovascular disease, cerebrovascular disease, nephropathy, neuropathy, retinopathy, and mortality. Using a variety of individual characteristics (sex, age, smoking status, body-mass index, etc.) that is updated annually, the MMD calculates transition probabilities for every individual into each of the different disease sub-states in each year. The MMD also tracks direct health care costs as well as quality of life estimates in every cycle of the simulation (1 cycle = 1 year), along with each individual's disease state. The quality-of-life index used in the MMD is the self-administered Quality of Well-Being (QWB-SA) scale [22]. The health utility values in the MMD are based on studies of diabetics using the QWB-SA to value health-related quality-of-life. Overall, this model simulates the progression of diabetes for a cohort of individuals and allows users to evaluate rates of complication development, health care costs, and utility of life estimates under user-defined circumstances [23-26].

We modify the baseline MMD by predicting depression status for each individual in every cycle. This prediction model was based on data from the RAND Health and Retirement Survey dataset (2000-2012) as well as the National Health and Nutrition Examination Survey (2007-2008). Based on sex, body-mass index, duration of diabetes, hypertension status, and

history of stroke and cardiovascular disease, we are able to predict with moderate discrimination individuals who should and should not have depression in a given year. Once an individual is assigned a positive depression status, they will experience two main changes in the model: increases in the probability of transitioning into the first disease state of each of the modeled comorbidities, and a depression-associated utility decrement of 0.10 for each corresponding cycle [27]. The first modification accounts for the documented increased risk of developing the comorbidities and complications associated with diabetes for patients with depression. Secondly, studies have identified an independent utility loss associated with depression that is not mediated through the increased risk of other diseases.

As mentioned earlier, individuals with depression also are characterized by increased health care costs and higher risks of mortality. These increases are not independently associated with depression however. The increased health care costs seem to be driven by non-mental health care costs, and thus are attributed to the increased risks of complication development for individuals with depression and diabetes. In this sub-population, the development of depression is associated with decreased treatment adherence, decreased diet, and decreased exercise, all of which can lead to multiplicative effects on adverse health outcomes and costs [10]. Similarly, the evidence on the independent risk of mortality is mixed thus far, with some research documenting an increased risk when controlling for individual characteristics and number of complications [28], while other research shows that controlling for the microvascular and macrovascular complications accounts for any significant relationship between depression and all-cause mortality [29]. Therefore, we do not change these two parameters in our modifications to the MMD, as we want to avoid double-counting or biasing our estimates of health care costs or mortality. We ran preliminary simulations comparing the mortality rates and costs between

depressed and non-depressed cohorts, and the depressed population experiences mortality rates up to 1.86 times as high as the non-depressed population, and cost increases that match the ranges reported in the literature. Therefore, we are confident that without an independent increase in the mortality rate or costs, the greater transition probabilities of the diabetes complications result in the characteristic increase in risk of death and higher health care costs. To summarize, we built a simulation model that predicts depression status of everyone in a simulation cohort and adjusts their probabilities of future complications as well as their associated utility scores accordingly. To use modeling to study the cost-effectiveness of the collaborative care intervention, we then had to develop treatment parameters for the usual care and intervention arms of our study. The usual care arm of most of the studies discussed earlier serves as our comparator/control group. This usual care consists of standard care by primary care physicians, with no extra emphasis on treating depression. Physicians can still administer pharmacotherapy, or refer patients to psychiatrists, but there is no multidisciplinary team monitoring patient progress and adjusting treatment as needed. We derive our estimates of the treatment effectiveness as well as the costs associated with usual care from the existing literature. Studies suggest that a usual care approach can be successful in approximately 30-40% of individuals who have depression [30-32]. Since there are no precise ways to predict which individuals will respond positively to the usual care treatment, the probability of successful treatment for individuals is equal for all individuals. As there may be considerable variation in the treatment effectiveness of usual care, depending on population or primary care physician, we vary this parameter in our sensitivity analyses from 20% – 60%, while the base-case effectiveness is 40%. From the literature, the estimated costs associated with administering usual care is approximately $402 per year [33]. Finally, in terms of quality of life effects, individuals

who are modeled as having been successfully treated in any given cycle do not incur any of the depression-associated utility decrement. We update all costs in our analyses to March 2016 US Dollars by using the Consumer Price Index.

Intervention

From the intervention standpoint, we model the effectiveness of the collaborative care approach that integrates providers, care managers, and patients with proactive monitoring and case review. The stepped care intervention is based out of primary care, and involves a care manager administering 1) pharmacotherapy and/or 2) problem-solving therapy and 3) frequent follow-up and active monitoring. This intervention takes place over 12 months only, with gradual decreases in the frequency of monitoring. For our analyses, we use results from the existing studies to inform our estimates of intervention effectiveness. In our specification of intervention effectiveness, we parameterize it as a benefit of treatment that is in addition to the usual care treatment. We used a standardized mean difference estimate from the literature as the additional effectiveness of the intervention when compared to usual care [5]. Therefore, if the base-case usual care effectiveness is 40%, the intervention effectiveness will be 40% plus the incremental effectiveness of the intervention at a given time. Unfortunately, the literature only provides estimates of collaborative care effectiveness up to 60 months. Therefore, we have to generate different effectiveness waning functions to develop a range of possible effectiveness estimates from year 1 to year 15 for our analyses.

These various waning functions rely on different time periods of the data available because the available data for the collaborative care intervention effectiveness is from studies among patients with both diabetes and depression (0 to 2 years follow-up) or just depression (0 to 5 years follow-up). So we use 0-24 month and 0-60-month data as the primary inputs for our

various waning functions. The waning function estimates we develop are: a) constant proportional waning rate (0-2 years), b) logarithmic best-fit waning function (0-5 years), c) constant waning, and d) a constant proportional waning rate (0-5 years). These waning functions use the existing effectiveness data as initiating data points, but then extrapolate into the long-term.

For the constant proportional waning rate from 0-2 years, we use the 2-year effectiveness data of the collaborative care intervention. Then, we assume that the proportion of patients who remain depression-free in year 2 compared to year 1 is a constant, and apply this to the subsequent years. In the logarithmic best-fit waning function, we take the data points of effectiveness in the first five years and fit a logarithmic function to the data. We then use the resulting function to estimate effectiveness for all time points. For the constant waning function, we assume that the absolute decline in effectiveness over the 5 years is constant, and apply this constant decrease to subsequent years. Finally, for the constant proportional waning rate (0-5 years), we use the data points from 5 years of effectiveness data and calculate the proportion of individuals that need to remain depression-free to match the effectiveness estimate at year 5. This assumes that each year a constant proportion of individuals will relapse.

Panel A of Figure 4.1 shows the different effectiveness estimates from these 4 methods over 15 years, while Panel B shows the average, minimum, and maximum effectiveness estimates for every year of our simulations. The values in these figures are the incremental improvements in effectiveness that are in addition to the usual care effectiveness. As seen in Panel B, the range of effectiveness we obtain from our waning functions covers the possible effectiveness rates, as we reach the floor of 0% effectiveness and a level slightly lower than the initial effectiveness for the minimum and maximum respectively.

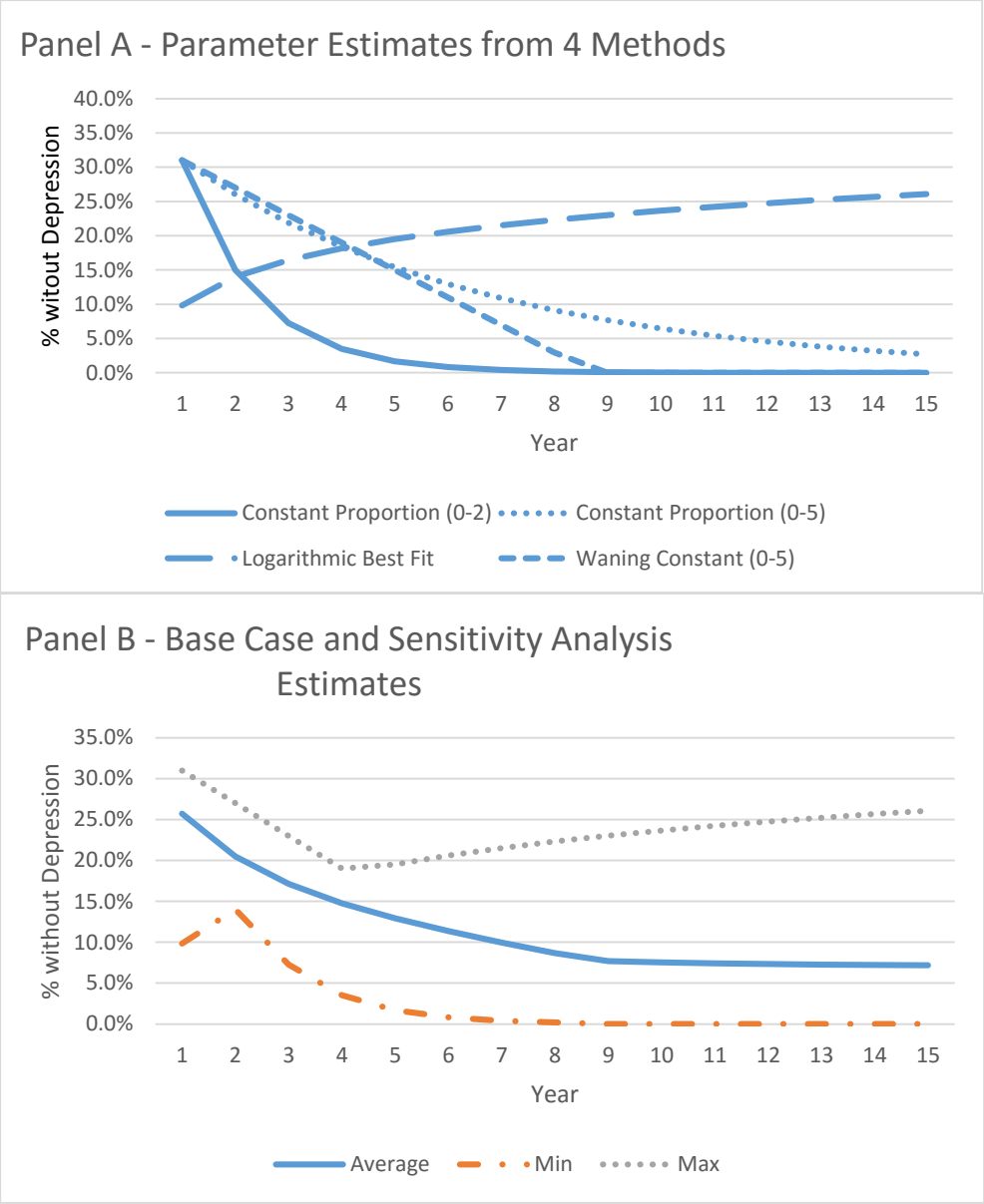


Figure 4.1 - Intervention Effectiveness Waning

In our parameterization, the intervention effectiveness is specified as the intervention effect in addition to usual care, so if there is complete waning of the intervention effect, the usual care treatment effectiveness will still apply. As there is no effectiveness data for the collaborative care treatment method beyond 5 years, these estimates of waning in effectiveness provide plausible parameters that can be used in our model for years 1-15. The ranges of effectiveness

data in each year will be used in sensitivity analyses so we can understand the impact of long-term effectiveness on the estimated cost-effectiveness.

For our intervention simulations, we assume a one-time collaborative care intervention that is administered to all individuals who have predicted depression in year 1. Then, in subsequent cycles, depression status is predicted at the beginning of each cycle, independent of depression status in preceding cycles. While we do not model the administration of the intervention again, we assume that there are incremental benefits of having experienced this intervention in year one that can lead to incrementally beneficial effectiveness compared to usual care. We model the intervention in this way because it allows individuals to continue receiving care after the intervention, without imposing an unrealistic end to usual care. Furthermore, the literature suggests this increased benefit lasts beyond the first year. Those in the intervention arm experience one-time intervention costs as well as usual care treatment costs whenever necessary. In our usual care simulations, all individuals who have predicted depression are assigned to usual care treatment irrespective of the cycle.

We are interested in both the short and long-term cost-effectiveness of the collaborative care treatment. Therefore, we calculate incremental cost-effectiveness ratios at 1, 5, 10, and 15 years. Intervention effectiveness dictates the transition probabilities an individual will experience throughout the Michigan Model for Diabetes, as those with depression after treatment have an increase in their risk of developing the modeled comorbidities compared to those individuals who become depression-free. Similarly, the individuals without depression experience a small decrease in their transition probabilities to match the literature estimates of the relative differences in transition probabilities between the two subpopulations. These different transition

probabilities will subsequently lead to heterogeneity in health care costs, and quality of life over many cycles.

The structure of the MMD requires that we run separate simulations for the intervention and control arms of our study (2 streams of Figure 4.2). In both the intervention and usual care simulations, we used predicted depression status in period 1 to identify which patients are eligible to receive treatment, while all non-depressed individuals in the cohort are removed from the simulation. Only the subset of individuals with depression in period 1 is subject to the treatment parameters throughout the rest of the simulation cycles. In the control simulation for example, once an individual is identified as having depression in period 1, they have a 40% probability of successful treatment. This treatment effect is instantaneous, so that those individuals who were predicted to have depression but then were successfully treated do not experience any increase in their transition probabilities or a decrement in utility. In subsequent cycles, depression status is again predicted based on the aforementioned algorithm, remaining functionally independent from previous depression status. Again, those who are identified as depressed at the beginning of each cycle are subject to the treatment effectiveness parameters associated with usual care (base-case: 40%). For the intervention simulations, the process is the same, but the intervention effectiveness parameter is specified as a benefit in addition to the usual care effectiveness.

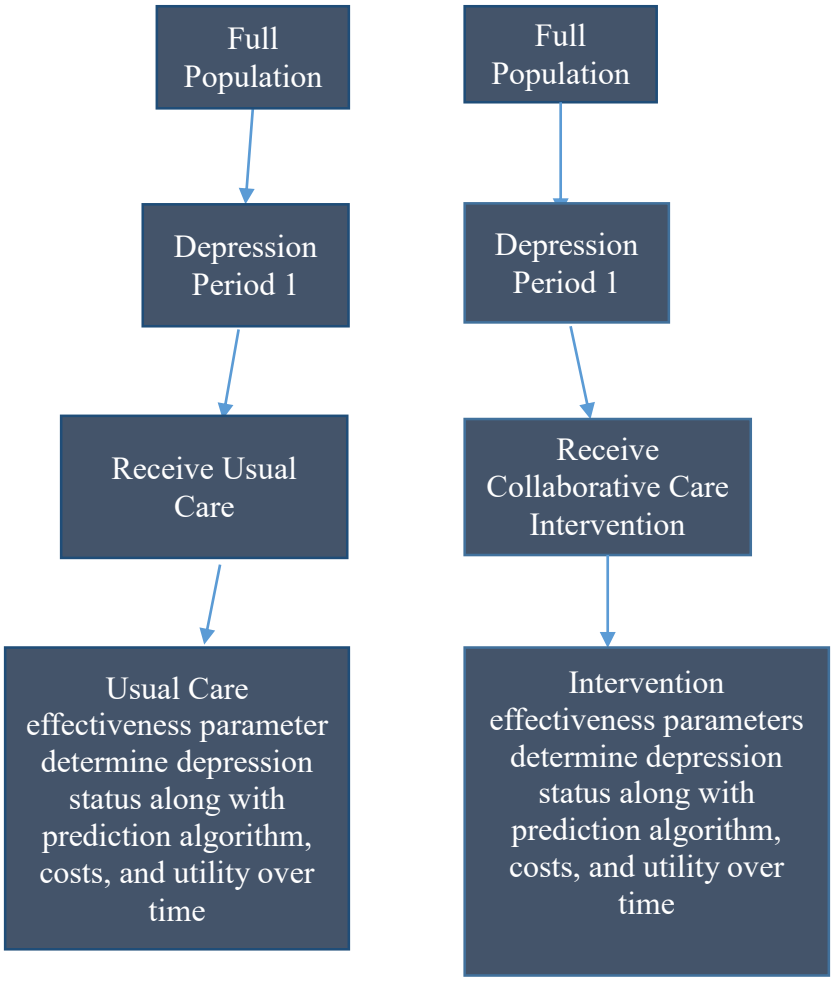


Figure 4.2 - Model Pathways

For those individuals with depression in period 10 for example, approximately 48% of individuals will be depression free in the intervention arm, while approximately 40% should be depression free in the usual care arm (using data from Figure 4.1).

For the population that does not have depression in period 1, since they would not receive any type of depression treatment in period 1, we remove them from the cohort. This helps align our analytic population with the trials done thus far, as treatment is only administered to individuals who start with depression. The simplifying assumption we are making here is that for those individuals who receive the intervention in period 1, they still should be able to develop

depression in future cycles because of the nature of this mental illness, but because they went through that intervention, their probability of successful treatment is higher than usual care (though this incremental benefit decreases over time). In the usual care simulations, individuals without depression in period 1 are removed, and then those who have depression receive the usual care treatment. From period 2 onwards, depression status is first predicted by the depression status algorithm, after which the usual care effectiveness randomly assigns which individuals are successfully treated and which ones have depression for that cycle. An important note is that in our model specification, effectiveness of treatment by both the collaborative care intervention and the usual care pathway is specified as a probability of success that is drawn from a Bernoulli distribution.

Cost and Quality Adjustments

The most relevant inputs for our model specification are seen in Table 4.1. Using the same study that informed our usual care cost estimates, our one-time intervention cost estimate is $703 for our model implementation [33]. The costs of intervention incorporate time spent on patient contacts, mean salary and benefits of care managers plus overhead costs, costs of supervision by psychiatrists and primary care experts plus overhead costs, and educational material costs [33]. We use the range of $532 - $1,264 as the intervention implementation cost estimates in our sensitivity analyses as reported in the other studies [16, 19, 20]. This is an incremental cost to the usual care treatment approach that everyone receives. For the usual care setting, our base-case input is $474, and the range we use in sensitivity analyses is $400-$600. The perspective of costs we use is the health care payer. This is the primary perspective of interest because in the administration of a collaborative care intervention, the payer is the one that incurs the direct costs. Furthermore, since the MMD captures direct health care costs, and

does not capture indirect costs, the health care payer perspective allows us to use our model without making assumptions about indirect cost effects. The utility loss associated with depression varies from -0.05 to -0.16, with our base-case estimate set at -0.10. Our estimates of the collaborative care intervention effectiveness are in Table 4.2, these incorporate the usual care effectiveness as well as the incremental collaborative care effectiveness. In our analyses, we use a 3% discount rate for all simulations. The costs in Table 4.1 are shown in 2014 US Dollars to match the rest of the MMD cost inputs. After each simulation, we then update all cost estimates to 2016 US Dollars using the Consumer Price Index to account for inflation.

Table 4.1 - Model Inputs

Parameters	***Inputs***
Utility Decrement	
Depression	-0.10 (-0.05 to -0.16)
Treatment Costs ($)	
Usual Care Costs per year	474 (400-600)
Collaborative Care (one-time)	703 (532-1264)

Table 4.2 - Depression Status After Intervention

	Depression Status			***Depression Status***	
Time	***0***	***1***	***Time***	***0***	***1***
1	65.71%	34.29%	8	48.66%	51.34%
2	60.51%	39.49%	9	47.70%	52.30%
3	57.14%	42.86%	10	47.54%	52.46%
4	54.76%	45.24%	11	47.42%	52.58%
5	52.91%	47.09%	12	47.33%	52.67%
6	51.34%	48.66%	13	47.27%	52.73%
7	49.95%	50.05%	14	47.22%	52.78%
			15	47.20%	52.80%

Cost-Effectiveness Analysis

Cost-effectiveness analysis allows for the comparison of costs and outcomes across interventions to improve health so that we can identify options that have the most ideal tradeoff between costs and outcomes [34]. While the numerator of the cost-effectiveness ratio is simply the difference in discounted aggregated costs between the two interventions, the denominator captures the difference in the aggregated discounted quality-adjusted life years (QALYs). The QALY provides a quantification of the gains from reduced morbidity and reduced mortality in a single measure [35]. Differences in QALYs across treatment options captures the expected benefits of a superior intervention [35]. The Quality of Well-Being Scale is a preference-weighted measure that incorporates morbidity and mortality to provide a numerical expression of well-being by assigning [36]. The QWB score provides a utility weight for the morbidity of a given health state, so that we can derive estimates of QALYS. Dividing the difference in costs by the difference in QALYs between an intervention and its comparator results in the incremental cost-effectiveness ratio (ICER):

$$ICER = \frac{(C_{Intervention} - C_{Control})}{(E_{Intervention} - E_{Control})} \quad where\ C = costs\ \&\ E = QALYs$$

When the ICER is below a willingness-to-pay threshold, the intervention is characterized as cost-effective. It is difficult to label a single value as the societal willingness to pay for improvements in health because different methods yield different values [37]. Reviewing the value-of-life literature, Hirth et al. find median $/QALY values ranging between $24,777 to $428,286 (in 1997 US Dollars) depending on the valuation method [38]. Although many cost-effectiveness analyses commonly cite a $50,000 or $100,000 per QALY threshold, evidence from the clinical setting suggests this threshold can be closer to the $200,000/QALY [39].

Recommendations now suggest using the $50,000 per QALY as a lower bound and that the WTP threshold can be thought to be ranging up to $200,000/QALY [37].

To calculate the incremental cost-effectiveness ratio of the collaborative care intervention at different time periods, we divide the difference in the average discounted total health care costs by the difference in the average discounted total utility scores [40]. We capture the total health care costs and total utility scores of all individuals who are alive in that time period, as well as the total health care costs and total utility scores of anyone that had died in a preceding time period. This ensures that we do not censor data from those who die during simulations, as their costs and utility estimates are valid inputs for the ICER calculation. All simulations are done using the Michigan Model for Diabetes after modifications to model depression and then data analyses were performed in Stata version 14.0.

We run simulations for 10,000 individuals over 15 years. Although our cohorts begin with 10,000 individuals in each simulation, since we remove all individuals who do not have depression in the first period, our analytic sample size is not the original 10,000. We use the same starting sample of 10,000 individuals in both the intervention and control arm to minimize population effects on estimates. The population cohort of the simulations is drawn from a distribution of user-generated inputs. The average age in our cohort population is 53, and the average duration of diabetes is 5 years. Finally, we use bootstrapping to quantify the uncertainty of these incremental cost-effectiveness ratio point estimates. Bootstrapping our sample 10,000 times allows for the calculation of confidence intervals around our incremental cost-effectiveness ratio, and provides us with an estimate of the willingness-to-pay above which we can be 95% confident that the intervention provides good value compared to the control simulations [40].

Model Assumptions

As with any modeling study, we make several assumptions to facilitate model development and fit existing computational constraints. An important one we make is that the treatment effectiveness of depression takes place at the beginning of each cycle. Predicted depression status is calculated first, but then before any changes to transition probabilities can take place, a proportion of those individuals will have gone through successful treatment so they are then deemed depression-free. We have to make this assumption, since depression status is recalculated at the beginning of every cycle. These individuals are treated successfully and they incur the cost of treatment but they do not have higher transition probabilities for any part of the cycle. This means that we miss any potential negative effects of having depression for part of the year before it is successfully treated by the end of the cycle. Secondly, since depression is a disease with high relapse rates, we do not allow successful treatment in preceding cycles to play a role in determining depression status at the beginning of future cycles. Therefore, the only determinants of depression status in each cycle are the depression prediction algorithm and the treatment effectiveness parameters, which are time-dependent in the intervention arm. Future work will investigate the time-dependency of depression treatment effectiveness to improve our model's clinical accuracy.

Although the intervention is only applied for one year, our assumption is that the individuals exposed to the collaborative care intervention have some added benefit in future years compared to control/usual care patients, and thus the persistence of an incremental effect in the base-case analysis is reasonable. In terms of costs, the treatment arm has a one-time cost of the collaborative care intervention, while after that individuals with predicted depression incur the control/usual care cost of treatment. In our control/usual care arm, we apply the effectiveness

parameter (40% in base-case analysis) to all patients with predicted depression in each cycle. Thus, we assume that there is no tapering of depression treatment effectiveness in the usual care scenario. It is imaginable that some patients will develop some type of resistance to the usual care treatment if it is continuously unsuccessful. Yet, since there are few data on the characteristics of patients that respond positively or negatively to depression treatment in both a cross-sectional and a longitudinal setting, we must make this simplifying assumption. Future work will address this limitation.

Sensitivity Analyses

As we make many assumptions regarding the base case parameter estimates we use in our model, robust sensitivity analyses are necessary to understand how the different variables may or may not impact the cost-effectiveness of the collaborative care treatment among individuals with diabetes and depression. In terms of effectiveness, we vary our estimates of the usual care as well as treatment effectiveness in one-way sensitivity analyses. The ranges of our waning functions cover the plausible range of effectiveness over the timeframe of our analyses, as it can range from complete waning in a couple of years, to a persistence of effectiveness. This helps determine how the difference between intervention and usual care effectiveness drives cost-effectiveness outcomes, and how the ICER may or may not change if the added benefit of treatment dissipates immediately.

Furthermore, the costs of treatment come from intervention implementation in fairly different settings. The administrative and overhead costs associated with implementation of this type of intervention could vary depending on location and health care provider organizations. Thus, varying this cost-estimate is important in improving the robustness of our findings. We also vary our control/usual care cost parameter, to see if higher or lower costs significantly

change the ICER. The evidence is limited as to the utility decrement associated with depression among diabetic patients when using the QWB-SA scale. To account for this parameter uncertainty, we run simulations varying this utility loss when individuals have depression to identify how sensitive the ICERs are to this value. Finally, we run a set of 2 and 3-way sensitivity analyses that model informative scenarios, including: high intervention effect & low intervention cost, and low intervention effect & high intervention cost. These scenarios will provide an estimate of the cost-effectiveness when the intervention is administered in the most optimal setting (lowest cost and greatest effect) or in the least optimal setting (highest cost and smallest effect). In our sensitivity analysis simulations, we model 10,000 individuals over 15 years, and use data from 1,000 bootstrap replications. The population for the sensitivity analyses is the same as the cohort population used in the base-case analyses.

Results

To first confirm that our model implementation is simulating through the cohorts in the intended manner, we calculated the proportion of individuals that ended each cycle with depression relative to the number identified as having depression in the beginning of each cycle. This process is necessary to make sure that our intervention and control models are successful in simulating depression treatment for all relevant individuals. Our treatment simulation always resulted in a higher proportion of individuals being depression-free, with the gap between treatment and control arms narrowing after the first 9-10 years (Figure 4.3). The values for the treatment simulation closely mirror the inputs in Table 4.2, while the consistency of the control simulation depression proportion being around 60% ensure that our model is simulating in the intended manner.

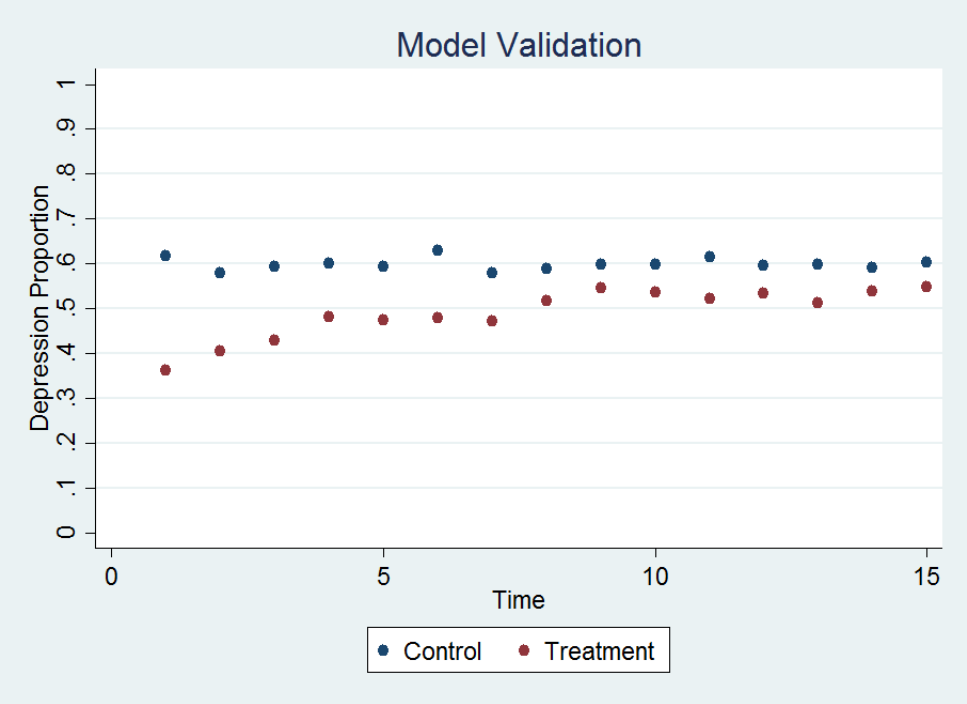


Figure 4.3- Model Validation

Table 4.3 provides a comparison of the intervention and usual care cohorts in terms of demographics and disease states across the different time points.

Table 4.3 - Demographic and Disease State Comparison Across Simulation Arms

Time	*Simulation*	*N*	*Mean Age*	*Male*	*BMI*	*Mortality*	*No Stroke*	*No CVD*	*No Nephropathy*	*No Neuropathy*
Period 1	Control	1986	53	30%	32	1.3%	90%	68%	84%	84%
	Intervention	1999	53	30%	32	1.4%	91%	68%	85%	83%
Period 5	Control	1894	57	28%	33	1.8%	91%	64%	67%	65%
	Intervention	1896	57	28%	33	1.4%	92%	64%	67%	67%
Period 10	Control	1750	62	25%	34	2.0%	93%	59%	51%	50%
	Intervention	1766	62	26%	35	1.5%	93%	58%	52%	52%
Period 15	Control	1573	66	22%	36	2.2%	94%	54%	38%	38%
	Intervention	1595	67	24%	36	1.9%	94%	55%	40%	38%

There was a relative level of balance in terms of the disease outcomes between the treatment and control groups between period 1 and period 15. Although it is expected that the control group should have greater rates of developing the comorbidities, since we attempted to model the relapsing nature of depression by recalculating depression status each year, the differences remain smaller in magnitude. If we had maintained the depression status of

individuals throughout multiple periods, the differences in disease states would have been larger in magnitude. Instead, we allow depression status and treatment effectiveness to be memoryless. Furthermore, in the later time periods, where differences in complication development may be more borne out, the differences in proportion of populations with depression will be less due to the waning of our treatment effectiveness. Finally, over time, the healthier individuals will remain in the sample. As seen in Table 4.3, the control group had a higher absolute mortality rate than the intervention group in periods 5, 10, and 15. This confirms that without an independent increase in the mortality risk, the higher probabilities of developing the complications and comorbidities translates into higher mortality rates in our simulations.

Turning towards the health care costs, health utility scores, and proportion with depression, Table 4.4 provides information on the aggregate discounted costs and utility scores across the simulations. In the first column, we list the discounted sum of all direct medical costs captured in the Michigan Model for Diabetes from the beginning of the simulation through the pertinent year. Similarly, the discounted total health utility score sums the yearly utility scores across all years. Finally, the depression column indicates the percentage of the simulation sample that had depression at that time period. These numbers are different than those seen in Figure 4.3, as here we just calculate the overall prevalence of depression. In Figure 4.3, we calculated the prevalence of depression among those individuals who started the cycle with depression and went through either control/usual care or intervention treatment.

Table 4.4 - Costs and Utilities by Time and Cohort

Time	*Simulation*	Discounted Total Cost	Discounted Total Health Utility Score	Depression
Period 1	Control	$ 7,971	0.50	62%
	Intervention	$ 8,635	0.52	36%
Period 5	Control	$ 36,516	2.40	38%
	Intervention	$ 35,675	2.47	29%
Period 10	Control	$ 69,772	4.42	38%
	Intervention	$ 67,644	4.52	34%
Period 15	Control	$ 97,770	6.11	43%
	Intervention	$ 95,992	6.23	39%

In Period 1, the treatment cohort cost $664 more on average than the control cohort. This estimate is close to the increased costs of the intervention group from trial data. In all subsequent periods, the total medical costs were higher in the control group. The aggregated discounted costs and utility values in each time period sum annual health care costs and utility scores for all individuals who started out in each cohort up to that year. Therefore, individuals who pass away in each simulation arm still contribute to the total cost and utility estimates. The discounted total health utility score was always higher for the treatment group, which should be driven in part by the higher proportion of individuals who avoid the decrement in health utility that is associated with depression. Based on Table 4.4 alone, we can predict that the incremental cost-effectiveness ratio point estimates will be favorable for the intervention due to the lower costs and higher utility. In our ICER calculations, we are interested in calculating the cost-effectiveness of the intervention up to the specified time period, so we use the total costs and total utility scores in our calculations to avoid biasing our estimates by removing individuals who die before the specified year.

Table 4.5 - ICER and Confidence Interval

	ICER	Quadrant	Bootstrap 95% CI
Period 1	27,469	Upper Right	Cost-Saving to 59,627
Period 5	Cost-saving	Lower Right	Cost-Saving to 35,641
Period 10	Cost-saving	Lower Right	Cost-Saving to 16,552
Period 15	Cost-saving	Lower Right	Cost-Saving to 25,920

The incremental cost-effectiveness ratio when comparing the collaborative care intervention to usual care to treat patients with both diabetes and depression was positive in the first year, and then cost-saving from period 5 onwards (Table 4.5). The bootstrap confidence intervals suggest that with a willingness-to-pay above $59,627, we would be 95% confident that

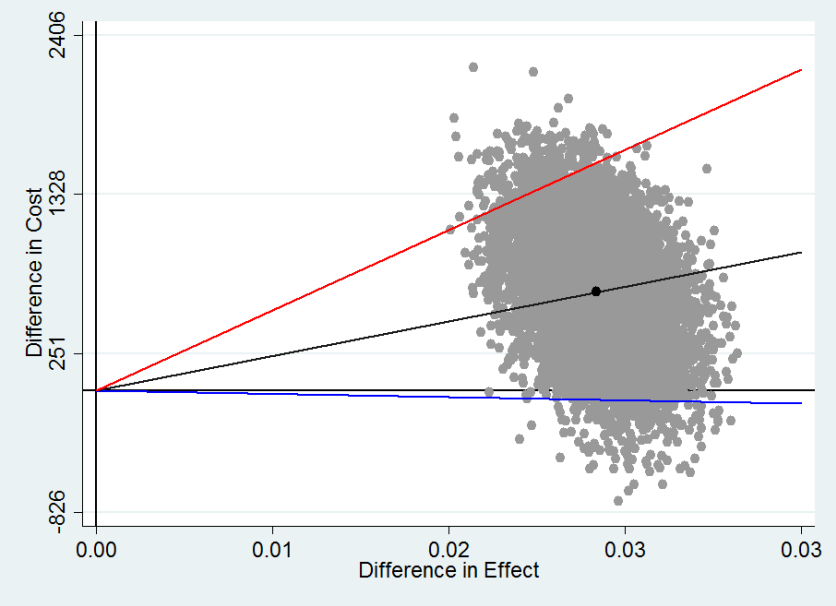


Figure 4.4 - Bootstrapped ICER in Period 1

the collaborative care intervention is significantly better value than usual care in all our studied time points. All of our ICER's after period 1 are in the lower right quadrant, which was expected due to the lower costs and increased benefits of the intervention arm.

Using data from the 10,000 bootstrap replications, we generated figures showing the results on the cost-effectiveness plane. The bootstrapped data for period 1 suggests that the collaborative care intervention will cost more money but have beneficial effects on QALYs (Figure 4.4). These figures show both the point-estimate (black dot), as well as the limits of the 95% confidence interval of the point-estimate (red and blue lines).

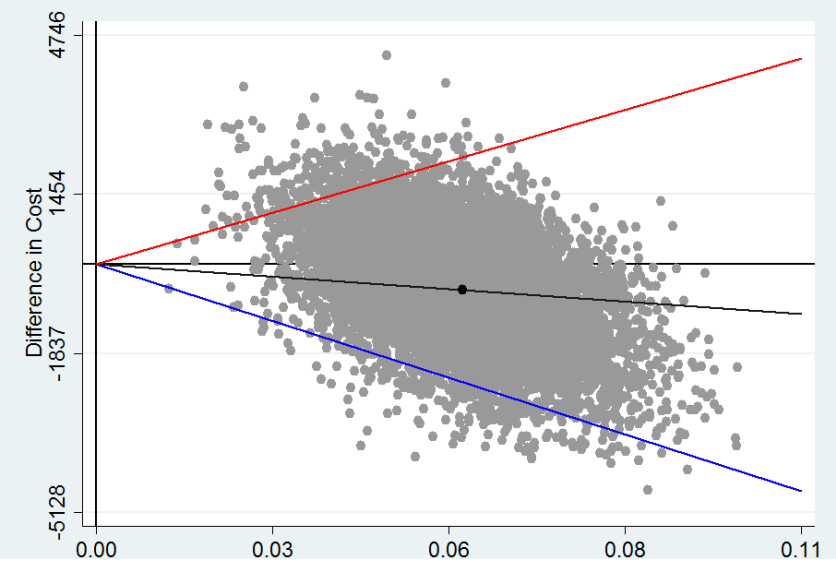


Figure 4.5 - Bootstrapped ICER in Period 5

Conversely, in Period 5, the point estimate is in the lower right quadrant, and many of the replications had cost-effectiveness ratios that were below the x-axis (Figure 4.5). When cost-effectiveness ratios cross the x-axis, they suggest that the intervention would be cost-saving compared to the control/usual care setting. This corresponds to the cost-saving lower bound of the confidence interval seen in Table 4.5.

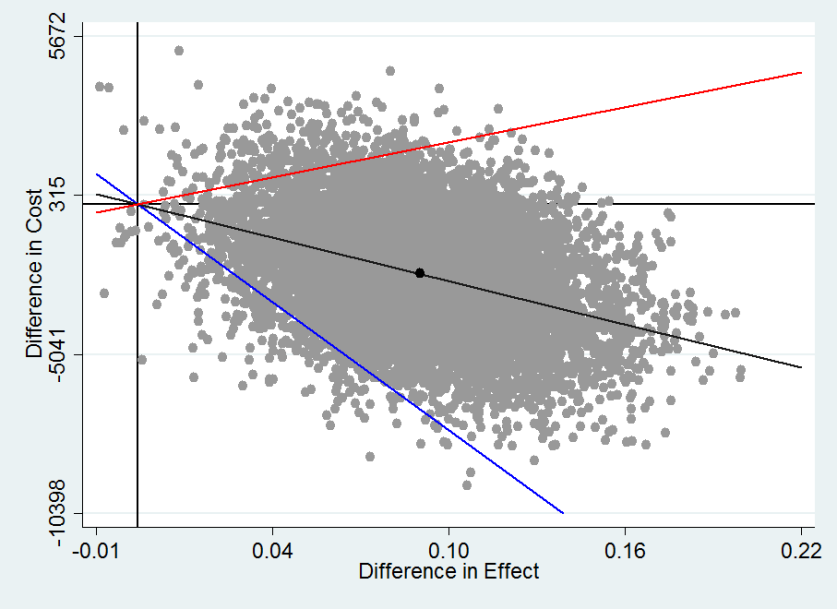


Figure 4.6 - Bootstrapped ICER Results in Period 10

This result was seen again for years 10 and 15, as all the ICER point-estimates as well as the majority of the bootstrapped replications were in the lower right quadrant (Figures 4.6 & 4.7). These findings suggest that in both the short-term (1-5) and the long-term (10-15 years), the collaborative care intervention can be a very high value investment as under our assumptions, as it is cost-saving from year 5 onwards.

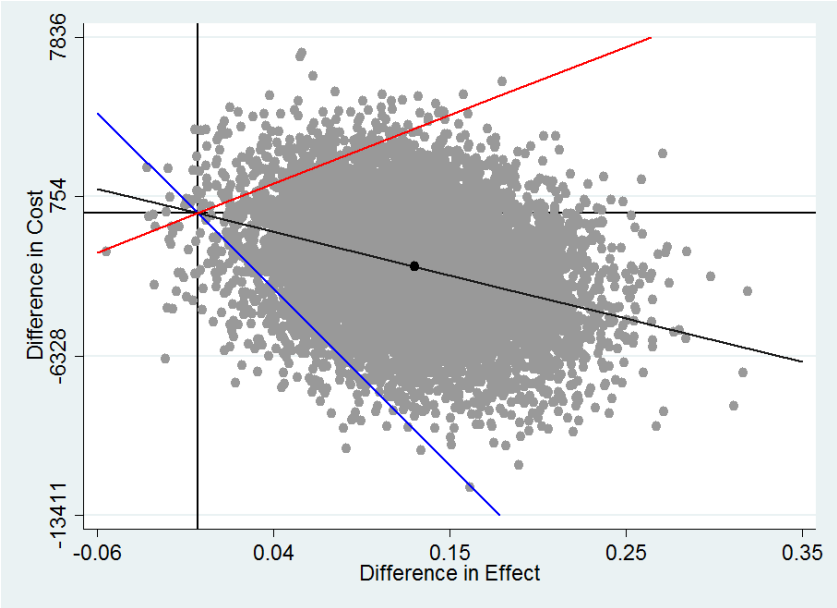


Figure 4.7 - Bootstrapped ICER Results in Period 15

These figures only show the possible range of incremental cost-effectiveness ratios, without incorporating willingness-to-pay. Accordingly, we generate cost-effectiveness acceptability curves that provide information on the probability the intervention can be considered cost-effective given a willingness-to-pay.

Above a WTP of approximately $70,595, the probability that the collaborative care intervention will be cost-effective compared to usual care reaches 99% in Period 1 (Figure 4.8).

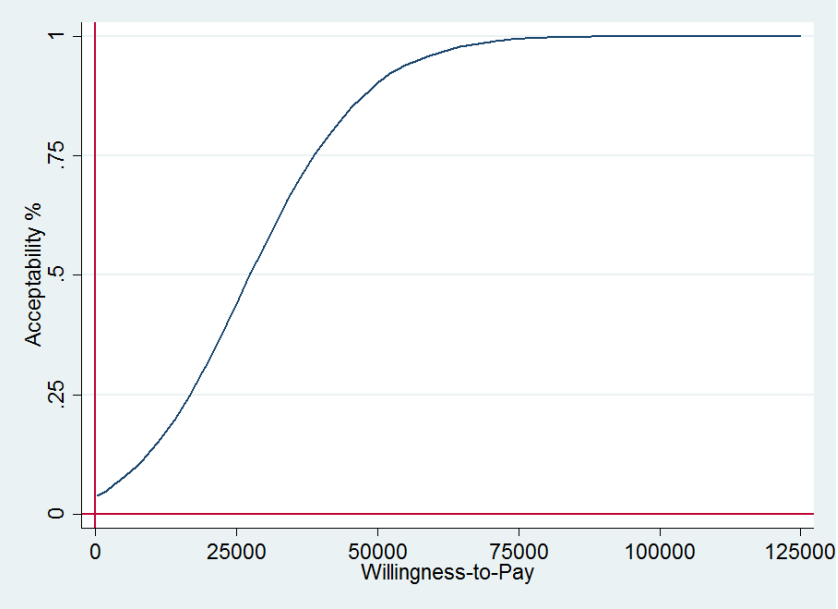


Figure 4.8 - Cost-Effectiveness Acceptability Curve in Period 1

When evaluating the cost-effectiveness acceptability results in year 15, we find results that correspond to Table 4.5, as a WTP above approximately $16,000 results in a greater than 95% probability of the intervention being cost-effective and good value compared to usual care (Figure 4.9). The cost-effectiveness acceptability curves for periods 5 and 10 are found in the Appendix (Figure 4A.1 & 4A.2). These figures show that the intervention would be cost-effective under commonly accepted willingness-to-pay values of $50,000-$200,000 per QALY [37]. It is important to note however, that these results are from our base-case analyses with the assumptions outlined earlier.

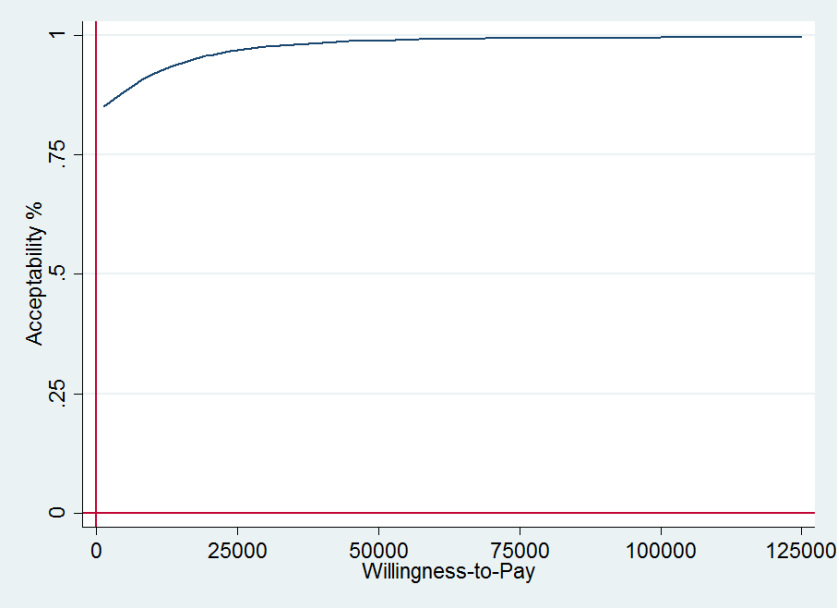


Figure 4.9 - Cost-Effectiveness Acceptability Curve in Period 15

Sensitivity Analysis Results

This section is organized by the variables used for sensitivity analyses. Results shown are from simulations with 10,000 individuals over 15 years, and then use data from 1,000 bootstrap replications to inform the cost-effectiveness acceptability and incremental cost-effectiveness

curves. The 10,000 individuals in these simulations are the same across all sensitivity analyses, and are the same as the ones used in the base case analysis. Since most of the uncertainty regarding the cost-effectiveness of the collaborative care intervention seems to be restricted to the first couple of years in the base-case analysis, we focus our sensitivity analyses on results between year 1 and year 10.

Utility

The data surrounding the utility decrement associated with depression among diabetic patients when using the Self-Administered Quality of Well-Being (QWB-SA) index is limited. Thus, we run simulations where we vary the utility decrement associated with depression, once with a decrement of -0.05 (small utility loss), and another time with a decrement of -0.16 (large utility loss). The cost-effectiveness acceptability curves show the effect the quality-of-life decrement estimate has on the probability of the collaborative care intervention being considered cost-effective for a variety of willingness-to-pays.

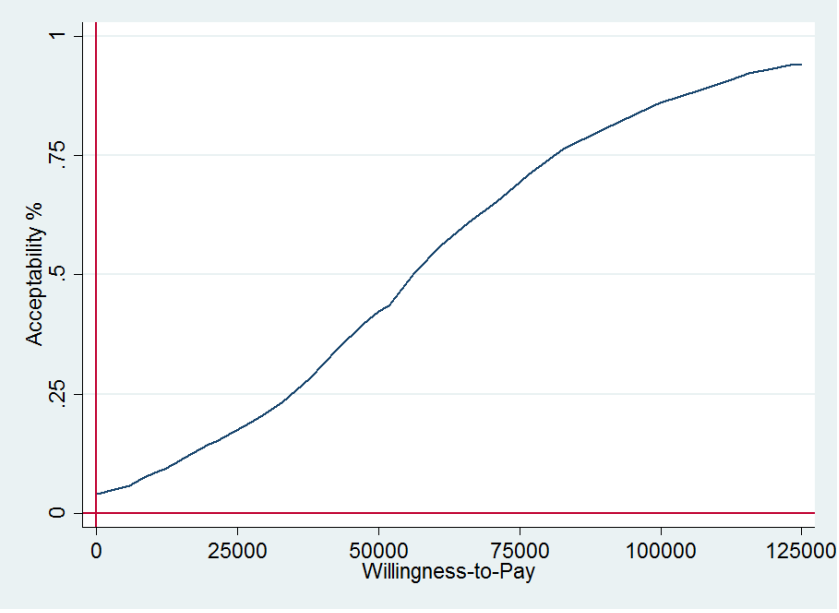


Figure 4.10 - Cost-Effectiveness Acceptability Curve (Utility -0.05, Period 1)

When the utility decrement associated with depression using the QWB-SA scale is reduced to only a -0.05 annual loss, the cost-effectiveness of the collaborative care intervention to treat patients with both depression and diabetes is greatly reduced (Figure 4.10). This is an expected result, as a lower utility decrement associated with a disease state reduces the benefical impact an intervention can have over time. A lower utility decrement reduces the average difference in utility scores between intervention and control simulations, shrinking the denominator and increasing the overall ICER. Conversely, if the utility loss associated with depression is increased in magnitude to -0.16, the intervention has as much more favorable cost-effectiveness profile (Figure 4.11). In period 10, by a WTP of $25,000, the intervention is highly likely to be cost-effective, while even at a lower willingness-to-pay, the probability never is below 50%.

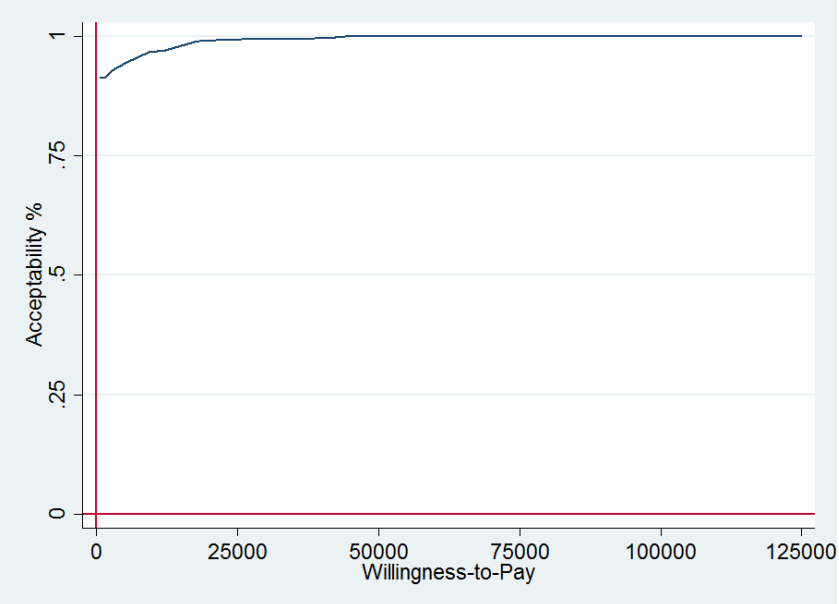


Figure 4.11 - Cost-Effectiveness Acceptability Curve (Utility -0.16, Period 10)

The differences between these two figures shows the effect of the depression-associated utility decrement on the cost-effectiveness of the collaborative care intervention approach. The

cost-effectiveness acceptability curves for period 1 are shown in the Appendix (Figures 4A.3, 4A.4).

We also ran analyses where we set the utility loss associated with depression to be 0. Conceptually, setting the independent utility loss from depression to 0 allows us to compare how much of the benefit of treating depression with the collaborative care intervention is from avoiding depression-specific utility decrements as opposed to avoiding utility losses from the increased risks of complications and comorbidities. In the base case analyses, the difference in total discounted utility between the control and treatment arms in year 15 was approximately 0.13 (5.52 vs. 5.65 respectively). When we remove the utility decrement associated with depression, the difference in total discounted utility between the control and treatment arms in year 15 was approximately 0.03 (5.97 vs. 6.00 respectively). While both scenarios resulted in collaborative care being dominant to usual care in year 15, the difference in utility losses reveals that approximately 23% of the benefit of the collaborative care treatment can be attributed to the effects on reduced complications. Conversely, 77% of the benefit of this treatment is attributable to the averted depression-associated utility loss.

Effectiveness

In our simulations, effectiveness is specified in two primary manners – intervention effectiveness and control/usual care effectiveness. As there is little evidence on the long-term effectiveness of the collaborative care intervention in a patient population with both diabetes and depression, varying this parameter is key to understanding how sensitive our ICER estimates may or may not be to this value. From Figure 4.1, we have lower and upper bound estimates for the effectiveness of the collaborative care intervention. The lower bound estimates allow us to calculate the cost-effectiveness of the collaborative care treatment approach when it only leads to

a miniscule marginal improvement in depression treatment compared to the usual care approach that tapers off by period 7. The upper bound estimates simulate the cost-effectiveness of the intervention if it was much more effective compared to the usual care, and maintained that larger marginal benefit for an extended period of time rather than waning quickly. When varying the baseline usual care effectiveness, we use a range of 20%-60%, to cover a wide range of possibilities in the effectiveness of standard care by primary care physicians when addressing the depressive symptoms of their patients.

The ICER point-estimates suggest relative intervention cost-effectiveness even when at the extremes of effectiveness (Table 4.6). Only when the intervention effectiveness parameter is at the worst does the period 1 incremental cost-effectiveness ratio increase slightly relative to baseline. While the point-estimates are similar to the base-case, closer inspection of the cost-effectiveness acceptability curves (Appendix Figure 4A.5) reveals that there is greater uncertainty surrounding the cost-effectiveness when treatment effectiveness is low. Even with this greater uncertainty, above a WTP of $100,000, the collaborative care intervention has a high probability of being cost-effective in the first year. In period 5, if the collaborative care intervention is highly ineffective, the intervention remains cost-effective above a WTP of $50,000 (Appendix Figure 4A.7 & 4A.9). Conversely, if the collaborative care intervention maintains a high rate of effectiveness for many periods when compared to continuous usual care, the cost-effectiveness point estimates are extremely favorable, remaining in either the lower right or upper right quadrants. The bootstrapped cost-effectiveness acceptability curves provide more evidence of the favorable cost-effectiveness profile (Appendix Figure 4A.6, Figure 4A.8).

Table 4.6 - ICER Dependence on Treatment Effectiveness

	Collaborative Care Treatment Effectiveness		Usual Care Effectiveness	
	Least Effective	*Most Effective*	*Least Effective*	*Most Effective*
Period 1	32,608	12,908	28,182	12,984
Period 5	Cost-Saving	Cost-Saving	26,050	Cost-Saving
Period 10	Cost-Saving	Cost-Saving	25,109	Cost-Saving

Note: All positive ICER's are in Quadrant I. LR – lower right quadrant

The incremental cost-effectiveness ratio is highly favorable when the usual care effectiveness is increased to 60%, as would be expected. For a small one-time intervention cost, the sustained reduction in depression rates would be highly cost-effective. When the usual care effectiveness is high, the collaborative care intervention is cost-saving, and when the usual care effectiveness is low, the ICER remains favorable, though this uncertainty is clearer from the cost-effectiveness acceptability curves (Figure 4A.12, Figure 4A.13, Figure 4A.14, Figure 4A.15).

Costs

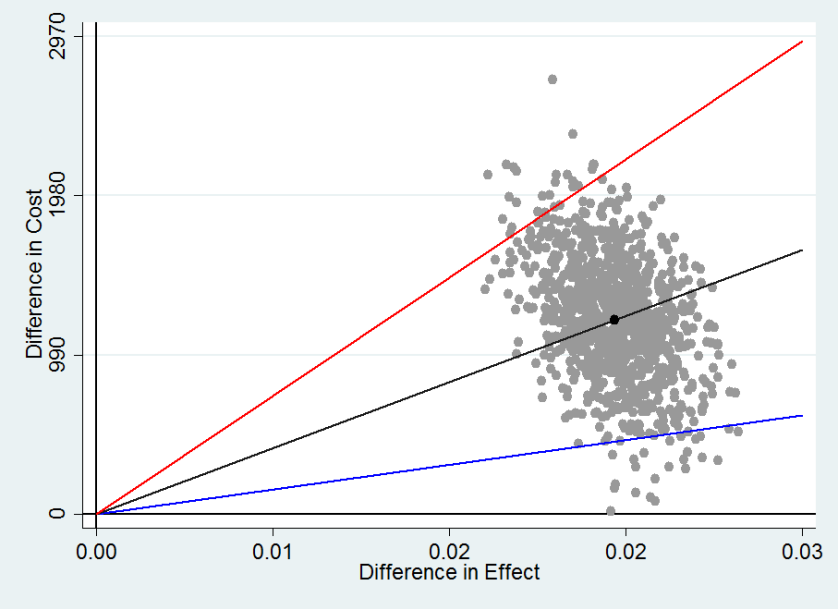


Figure 4.12 - Bootstrapped ICER (High Intervention Cost, Period 1)

We also evaluate how changes to the costs associated with both treatment strategies may or may not change the cost-effectiveness profile of the collaborative care intervention. Using data from the existing studies, we vary the one-time cost associated with the collaborative care treatment from $532 - $1,264. Data from 1,000 bootstraps suggests that the ICER is positive when the intervention costs $1,264 (Figure 4.12). Compared to Figure 4.10, we see that a higher willingness-to-pay is needed to be extremely confident that the collaborative care intervention is good value when the intervention cost is higher (Figure 4.13), but even using the suggested range of willingness-to-pay ($50,000-$200,000/QALY), the probability of the intervention being cost-effective is fairly high.

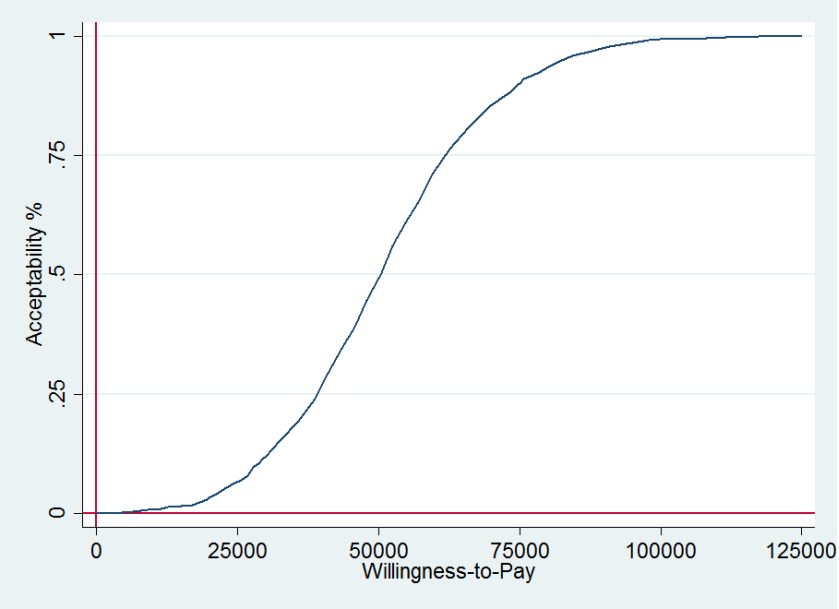


Figure 4.13 - Cost-Effectiveness Acceptability Curve (High Intervention Cost, Period 1)

In our simulations, the slight increase in the intervention cost does not seem to change the cost-effectiveness acceptability that much compared to the base-case analyses. This change in intervention cost is small in magnitude to the overall costs. When we set the one-time cost of the intervention to be $532, the cost-effectiveness profile of the treatment becomes even more

favorable compared to the base-case analyses. As seen below, the bootstrapped ICER point estimate (Figure 4.14) as well as the cost-effectiveness acceptability curve in period 1 (Figure 4.15) for this scenario show that the intervention would be cost-effective.

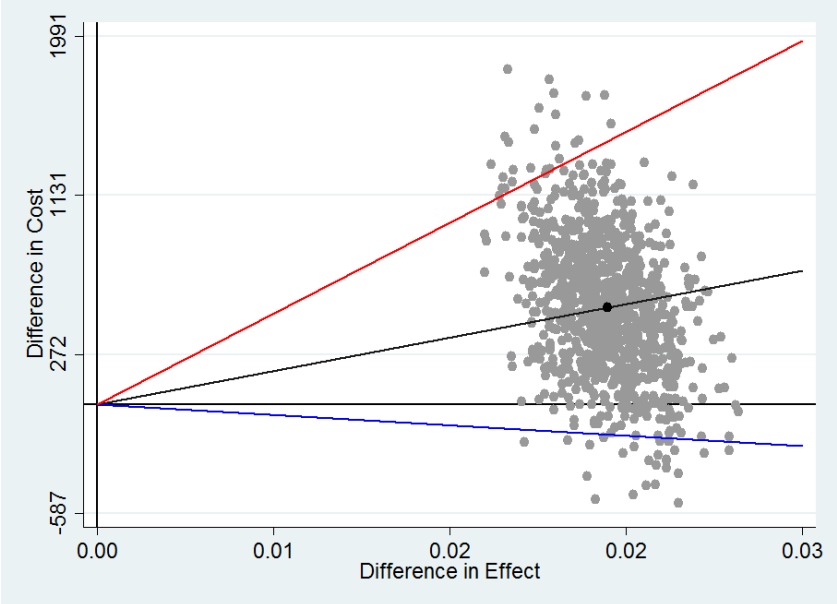


Figure 4.14 - Bootstrapped ICERs (Low Intervention Cost, Period 1)

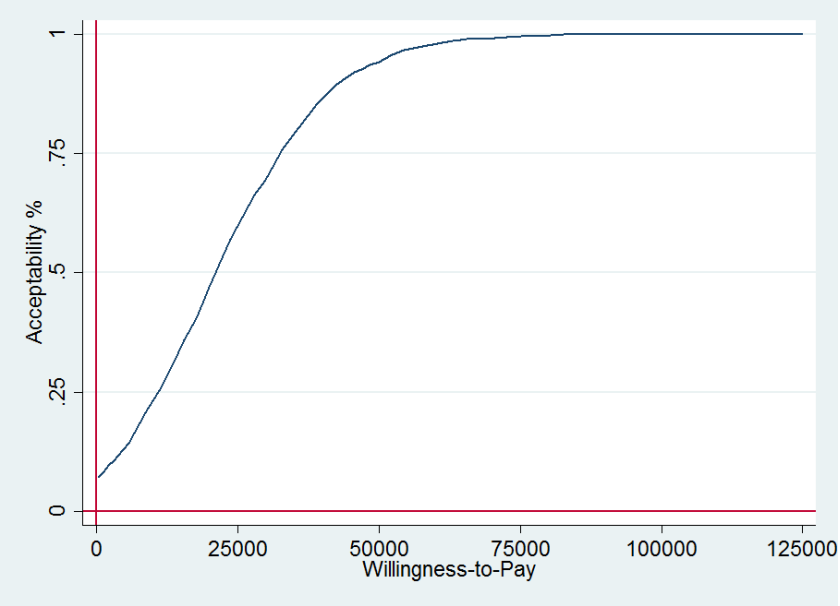


Figure 4.15 - Cost-Effectiveness Acceptability Curve (Low Intervention Cost, Period 1)

Therefore, as the costs associated with utilizing a collaborative care intervention decrease, the intervention becomes more dominant compared to the usual care treatment, holding all other parameters constant. Even when increasing the intervention costs, the intervention would remain cost-effective above relatively small willingness-to-pays.

When varying the usual care costs from $400 - $600, the cost-effectiveness acceptability curves maintain a relatively similar shape. This matches expectations, as the changes in the usual care cost would lead to a small level shift of the cost-effectiveness acceptability curve in either direction (Figure 4A.10, Figure 4A.11).

Two-Way Sensitivity Analyses

Simultaneously varying multiple variables of interest allows us to calculate the cost-effectiveness of the intervention under scenarios that simulate either worst- or best-case scenarios. We run two-way sensitivity analyses around the intervention effectiveness and intervention cost. For the collaborative care intervention, the least optimal scenario would be if the intervention was minimally effective and very costly. Under these settings, the cost-effectiveness of the collaborative care intervention in period 1 increases to $88,260. Conversely, if the intervention operates in the most ideal scenario with the lowest-associated cost and the greatest effectiveness, the period 1 cost-effectiveness ratio is $13,060. As evidenced by the comparison of the cost-effectiveness acceptability curves (Figures. 4.16 & 4.17), we see the variation in WTP needed between the most optimal and least optimal scenario.

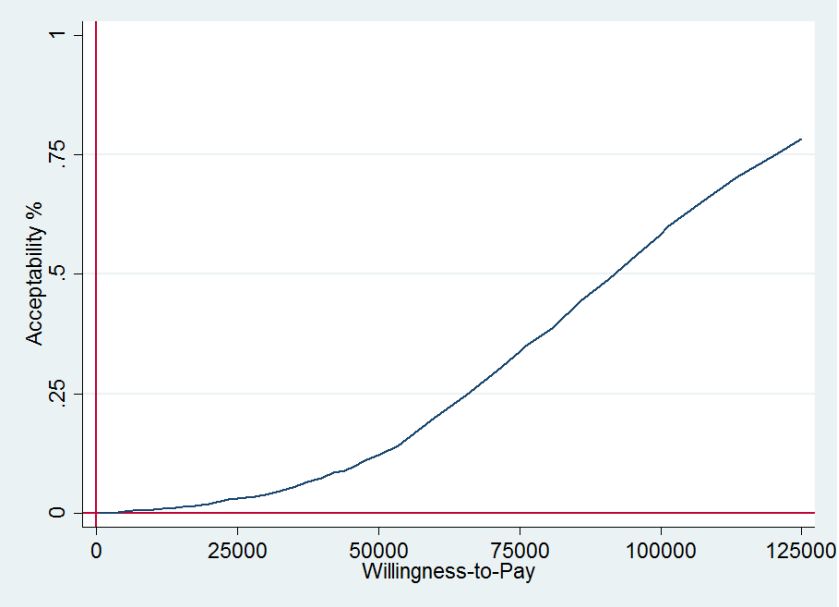


Figure 4.16 - Cost-Effectiveness Acceptability Curve (Low Effect & High Cost, Period 1)

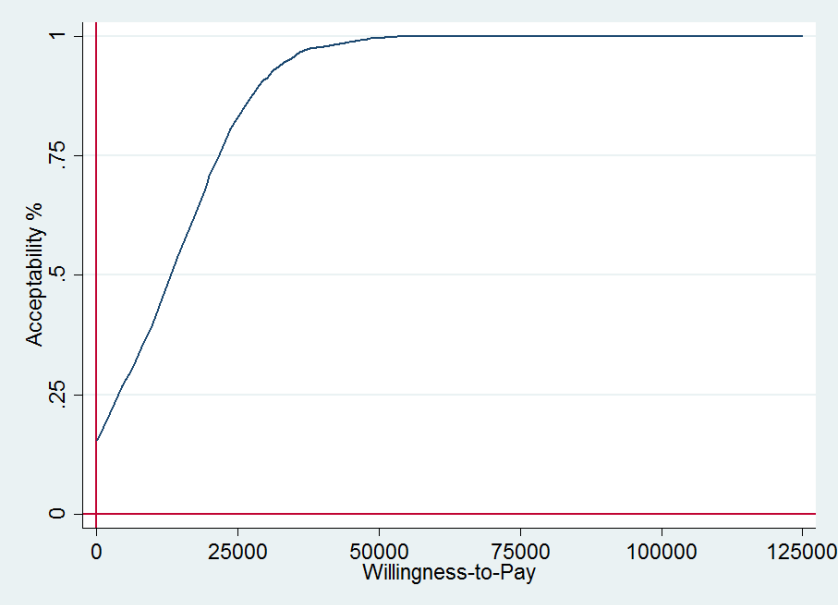


Figure 4. 17 - Cost-Effectiveness Acceptability Curve (High Effect & Low Cost, Period 1)

The three-way sensitivity analysis we perform simulates the collaborative care intervention under the worst scenario, where the intervention cost is the highest, the effectiveness is the lowest, and the utility decrement reduced to -0.05. In this scenario, the collaborative care intervention has a positive ICER in periods 1 and 5, but then becomes cost-saving in periods 10 and 15. The difference between the control and intervention groups in terms of costs and utility scores is much smaller in this scenario, but the differences are large enough to result in the intervention being cost-saving in the later periods.

Limitations

As with most modeling and cost-effectiveness studies, there are limitations to our analyses. From a cost perspective, the current version of the Michigan Model for Diabetes only captures direct medical costs of diabetes and the associated comorbidities and complications. Thus, indirect costs such as productivity costs, are not captured in our analyses. Further, the costs and utility losses/benefits associated with informal caregiving are another set of inputs that can inform cost-effectiveness analyses. Especially in the context of diabetes and depression, two illnesses that can have large care burdens on unpaid caregivers, including these costs in cost-effectiveness analyses would help results be more robust and complete [41, 42]. As the data is limited as to how the interaction of diabetes and depression would change the informal care cost estimates (whether it is additive or multiplicative), any inclusion of informal care costs would be limited to preliminary analyses. Further, there will be inherent heterogeneity in the informal costs for a patient with diabetes based on their characteristics and complication history. With more research into independent drivers of informal care costs among patients with diabetes and depression, we may be able to add this source of costs to our estimates.

Turning to our effectiveness parameters, the limited evidence surrounding the patient population with both diabetes and depression necessitates assumptions in model development. Our effectiveness parameters were developed using a variety of waning functions, and the range of collaborative care intervention effectiveness allows us to explore how the cost-effectiveness of this treatment may or may not vary depending on the success of treatment. Nevertheless, the uncertainty surrounding these parameters is a limitation. Additionally, the memoryless nature of our logic in determining depression status in each period has both advantages as well as disadvantages. Since depression is a very complex disease with a high rate of relapse, this memoryless component allows us to partially capture this complexity where past history doesn't monopolize future experiences with depression. Conversely, for an individual patient, their responsiveness to depression treatment may be driven to some degree by their history. Since there is limited longitudinal data on depression among patients with diabetes, we cannot capture these dynamics in the MMD yet. The issue of uncertainty in the waning of treatment effectiveness also is important in the usual care arm of our simulations. We assume that there is no waning of effectiveness at the individual level. As the evidence is limited in identifying which patients with diabetes respond to treatment, and what type of waning there may be for these individuals in treatment effectiveness over time, we make the simplifying assumption that the probability of response to treatment is equal across all patients over all years of our simulations. This bias probably leads to underestimates of the cost-effectiveness of the collaborative care intervention, as it presumably would be more likely that patients less responsive to general treatment may experience effective treatment with the collaborative care intervention. Subsequently, the reduction in complication risk for these patients would lead to larger differences in effects when comparing the collaborative care intervention with usual care.

This memoryless component may be a driving factor behind the limted differences in complication development (Table 4.3) between the two arms. If we had better data on the length of depressive symptoms and the predictors of length, we would have been able to model depression length instead of having to rely on the cycle length dictated by the model. So a more robust model would allow for heterogeneity in depression episode length, which could subsequently allow the differences in complication development to be more borne out even after incorporating a treatment parameter. If we modeled depression episode length, then patients with longer episodes of depression would have uninterrupted increases in their risks of future complications. Another consequence of our MMD modifications is that our analytic sample is disproportionately female. Although the prevalence of depression is significantly increased among females [7], our model may overemphasize this difference. Nevertheless, the results from our analyses provide robust base-case estimates of the cost-effectiveness of this intervention, and we plan to engage in future research to address some of these limitations.

Finally, we are unable to perform probabilistic sensitivity analyses through the Michigan Model for Diabetes, which would further inform the evidence surrounding the cost-effectiveness of the collaborative care intervention. Instead, we run a variety of sensitivity analyses on the parameters that seem to be the most salient for providers when they would consider using this treatment modality to address depression among patients with diabetes.

Discussion

The collaborative care treatment approach combines physicians, nurses, and patients in a systematic stepped care approach to address depression symptoms among patients. While previous studies have found this approach to be cost-effective when treating patients with both depression and diabetes, no long-term cost-effectiveness data are available to our knowledge.

Using a microsimulation model that allows users to monitor the progression of diabetes among a cohort of individuals, we model the development and treatment of depression. We derive our parameter inputs from the existing studies, and use various waning functions to predict the effectiveness of the intervention beyond 5 years. In our base-case analysis, we find that above a willingness-to-pay **of** $59,627 the intervention is cost-effective in the short- and long-term. While there are different probabilities of the intervention being cost-effective across willingness-to-pays and year, our data suggests that the collaborative care intervention can be very cost-effective. Under the recommended range of willingness-to-pay values between $50,000-$200,000 per QALY, the intervention would be considered cost-effective [37].

In our analyses, we aggregate direct medical costs for individuals in our simulations, as well as costs associated with the treatment of depression, and deduct a depression-associated utility decrement whenever necessary. Total direct medical costs were lower for the intervention simulations compared to the usual care simulation after year 1, while total health utility scores were higher in our intervention simulation arms. In our base-case analyses, the incremental cost-effectiveness ratio point-estimates were cost-saving for the collaborative care intervention in year 5, 10, and 15. Under suggested willingness-to-pay values, our evidence suggests that this intervention would be a good value investment.

Our sensitivity analyses show that if the utility loss associated with depression is small in magnitude, the intervention is less likely to be considered cost-effective. This is an expected result, since a small utility loss would limit the effectiveness gains any treatment could have. On the other hand, if the quality-of-life loss from depression is higher in magnitude than our base-case, the intervention cost-effectiveness profile improves drastically. When modeling the worst-case scenario with regards to the effectiveness of the collaborative care intervention, where the

incremental benefit compared to usual care treatment is minimal and is nonexistent by year 7, the intervention would require a larger willingness-to-pay to be considered cost-effective. Conversely, when we model the intervention as maintaining a high level of incremental effectiveness for multiple years, or various levels of usual care effectiveness, the intervention seems to be highly cost-effective. Thus, if there was a context where the collaborative care intervention has a very small impact on depressive symptoms, then the intervention may not be cost-effective. In all other instances where there is variance in the effectiveness of the intervention or the baseline standard care effectiveness, this intervention could be cost-effectiveness based on our model. As would be expected, a low one-time intervention cost improves the cost-effectiveness profile.

Therefore, the variables that seems to be most important in determining the cost-effectiveness of the collaborative care intervention under our assumptions is the utility decrement associated with depression and treatment effectiveness. If the independent effect of depression on health utility is small, then higher willingness-to-pays are necessary to improve the likelihood of the intervention being cost-effective. Secondly, if the intervention is very ineffective, where the incremental benefit compared to usual care never goes above 15% and only increases the effectiveness of treatment by 2% in period 5, a higher willingness-to-pay is needed to increase the likelihood of cost-effectiveness. On the other hand, the base-case analyses as well as increased utility decrements, increased intervention effectiveness, and a variety of intervention and usual care costs suggest that the intervention can be very cost-effective. Previous analyses have documented beneficial cost-effectiveness ratios of this treatment approach [17,19-21] , so our findings align with what has been previously reported.

Conclusions

The collaborative care approach to treating depression among patients is becoming more popular. As more health care institutions consider investing in using this type of interdisciplinary approach to address the needs of patients with depression, evidence is needed to make informed decisions. Using a modeling-based approach allows us to study the health and economic benefits of treatment strategies that have limited longer-term evidence. Under our assumptions of treatment effectiveness and cost, as well as standard usual care costs and effectiveness, we find evidence to suggest that the collaborative care intervention can be cost-effective under the recommended range of willingness-to-pays. In addition, we provide our evidence from a variety of sensitivity analyses so decision-makers with situation-specific knowledge of the costs, effects, or utility losses associated from depression may be able to develop a better idea of the expected cost-effectiveness of this treatment approach.

As screening of adults for depression increases to meet the United States Preventive Services Task Force updated recommendations [43], the diagnosed prevalence of depression among patients with diabetes may increase. For many years the underdiagnosed nature of depression among individuals with diabetes has hampered treatment efforts. With the increase in screening efforts, more individuals may be identified with depressive symptomology, and effective treatment methods will be needed to address the complex health care needs of individuals with depression. For patients with depression and diabetes, the collaborative care intervention based out of primary care seems to be a viable treatment modality with low associated costs and relatively benefical depression outcomes. Using a modeling approach, we find that the collaborative care intervention is high value with a very favorable cost-effectiveness profile.

References

1. Bodenheimer T, Chen E, Bennett HD. Confronting the growing burden of chronic disease: can the u.s. health care workforce do the job? Health Affairs, 2009. 28(1): p. 64-74.

2. Shaw JE, Sicree RA, and Zimmet PZ, Global estimates of the prevalence of diabetes for 2010 and 2030. Diabetes Research and Clinical Practice, 2010. 87(1): p. 4-14.

3. Anderson RJ, et al., The prevalence of comorbid depression in adults with diabetes: a meta-analysis. Diabetes Care, 2001. 24(6): p. 1069-78.

4. Mezuk B, et al., Depression and type 2 diabetes over the lifespan: a meta-analysis. Diabetes Care, 2008. 31(12): p. 2383-90.

5. Atlantis E, Fahey P, Foster J., Collaborative care for comorbid depression and diabetes: a systematic review and meta-analysis. BMJ Open, 2014. 4(4).

6. Katon WJ et al., Collaborative care for patients with depression and chronic illnesses. New England Journal of Medicine, 2010. 363(27): p. 2611-2620.

7. Anderson RJ, et al., The prevalence of comorbid depression in adults with diabetes. a meta-analysis, Diabetes Care 2001. 24(6): p. 1069-1078.

8. Black SA,Markides KS, Ray LA, Depression predicts increased incidence of adverse health outcomes in older Mexican Americans with type 2 diabetes. Diabetes Care, 2003. 26(10): p. 2822-8.

9. Katon WJ, et al. The association of comorbid depression with mortality in patients with type 2 diabetes. Diabetes Care, 2005. 28(11): p. 2668-72.

10. Ciechanowski PS, Katon WJ, Russo JE. Depression and diabetes: Impact of depressive symptoms on adherence, function, and costs. Archives of Internal Medicine, 2000. 160(21): p. 3278-3285.

11. Simon GE, et al., Diabetes complications and depression as predictors of health service costs. Gen Hosp Psychiatry, 2005. 27(5): p. 344-51.

12. Molosankwe I, et al., Economic aspects of the association between diabetes and depression: a systematic review. J Affect Disord, 2012. 142 Suppl: p. S42-55.

13. Beaulieu N, et al., The business case for diabetes disease management for managed care organizations. Forum for Health Economics & Policy. 2006.

14. Archer, J et al., Collaborative care for depression and anxiety problems. Cochrane Database Syst Rev, 2012. 10: p. Cd006525.

15. Gilbody S, et al., Collaborative care for depression: a cumulative meta-analysis and review of longer-term outcomes. Arch Intern Med, 2006. 166(21): p. 2314-21.

16. Simon,GE, et al., Cost-effectiveness of systematic depression treatment among people with diabetes mellitus. Arch Gen Psychiatry, 2007. 64(1): p. 65-72.

17. Katon WJ, et al., Cost-effectiveness of improving primary care treatment of late-life depression. Arch Gen Psychiatry, 2005. 62(12): p. 1313-20.

18. Katon WJ, et al., Cost-effectiveness and net benefit of enhanced treatment of depression for older adults with diabetes and depression. Diabetes Care, 2006. 29(2): p. 265-270.

19. Katon WJ, et al., Cost-effectiveness of a multicondition collaborative care intervention: a randomized controlled trial. Arch Gen Psychiatry, 2012. 69(5): p. 506-14.

20. Hay JW, et al., Cost-effectiveness analysis of collaborative care management of major depression among low-income, predominantly Hispanics with diabetes. Value Health, 2012. 15(2): p. 249-54.

21. Jeeva F, et al., Is treatment of depression cost-effective in people with diabetes? A systematic review of the economic evidence. Int J Technol Assess Health Care, 2013. 29(4): p. 384-91.

22. Coffey JT, et al., Valuing health-related quality of life in diabetes. Diabetes Care, 2002. 25(12): p. 2238-2243.

23. Herman WH, et al., Early detection and treatment of type 2 diabetes reduce cardiovascular morbidity and mortality: a simulation of the results of the anglo-danish-dutch study of intensive treatment in people with screen-detected diabetes in primary care (addition-europe). Diabetes Care, 2015.

24. Zhou H, et al., A computer simulation model of diabetes progression, quality of life, and cost. Diabetes Care, 2005. 28(12): p. 2856-63.

25. Isaman DJ, Barhak J, Ye W, Indirect estimation of a discrete-state discrete-time model using secondary data analysis of regression data. Stat Med, 2009. 28(16): p. 2095-115.

26. Barhak J, et al., Chronic disease modeling and simulation software. J Biomed Inform, 2010. 43(5): p. 791-9.

27. Pyne JM, et al., Use of the quality of well-being self-administered version (QWB-SA) in assessing health-related quality of life in depressed patients. J Affect Disord, 2003. 76(1-3): p. 237-47.

28. Katon WJ, et al., The association of comorbid depression with mortality in patients with type 2 diabetes. Diabetes Care, 2005. 28(11): p. 2668-2672.

29. Bruce DG, et al., A prospective study of depression and mortality in patients with type 2 diabetes: the Fremantle Diabetes Study. Diabetologia, 2005. 48(12): p. 2532-2539.

30. Johnson JA, et al., Collaborative care versus screening and follow-up for patients with diabetes and depressive symptoms: results of a primary care-based comparative effectiveness trial. Diabetes Care, 2014. 37(12): p. 3220-6.

31. Bogner HR., et al., Integrated management of type 2 diabetes mellitus and depression treatment to improve medication adherence: a randomized controlled trial. Ann Fam Med, 2012. 10(1): p. 15-22.

32. Unutzer J et al., Collaborative care management of late-life depression in the primary care setting: a randomized controlled trial. Jama, 2002. 288(22): p. 2836-45.

33. Katon W, et al., Cost-effectiveness and net benefit of enhanced treatment of depression for older adults with diabetes and depression. Diabetes Care, 2006. 29(2): p. 265-70.

34. Gold MR, Siegel JE, Russell LB, Weinstein MC. (1996). *Cost-effectiveness in health and medicine*. Oxford, UK: Oxford University Press

35. Drummond M, Sculper MJ, Torrance G, O'Brien BJ, & Stoddard, GL. (2005). *Methods for the economic evaluation of health care programmes* (3rd ed.). Oxford, UK: Oxford University Press.

36. Kaplan RM, et al., The Quality of Well-Being Scale: critical similarities and differences with SF-36. Int J Qual Health Care, 1998. 10(6): p. 509-20.

37. Neumann PJ, Cohen JT, Weinstein MC. Updating cost-effectiveness – the curious resilience of the $50,000-per-QALY threshold. New England Journal of Medicine, 2014. 371: 796-797.

38. Hirth RA, Chernew ME, Miller E, Fendrick AM, Weissert WG. Willingness to pay for a quality-adjusted life year: in search of a standard. Med Decis Making, 2000. 20(3): 332-342.

39. Ubel PA, Hirth RA, Chernew ME, Fendrick AM. What is the price of life and why doesn't it increase at the rate of inflation? Arch Intern Med, 2003. 163(14): 1637-1641,

40. Polsky D, et al., Confidence intervals for cost-effectiveness ratios: a comparison of four methods. Health Econ, 1997. 6(3): p. 243-52.

41. Langa KM, et al., Informal caregiving for diabetes and diabetic complications among elderly americans. The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 2002. 57(3): p. S177-S186.

42. Bock JO, et al., Impact of depression on health care utilization and costs among multimorbid patients – results from the multicare cohort study. PLoS One, 2014. 9(3).

43. Siu AL., and U.S.P.S.T.F. and the, Screening for depression in adults: Us preventive services task force recommendation statement. JAMA, 2016. 315(4): p. 380-387.

Appendix

Base Case Cost-Effectiveness Acceptability Curves

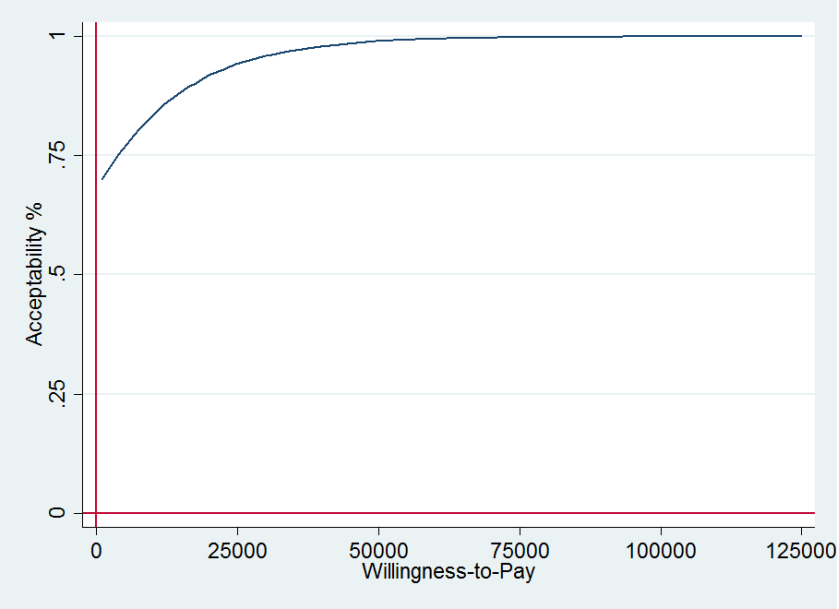


Figure 4A.1 - Cost-Effectiveness Acceptability Curve (Base Case, Period 5)

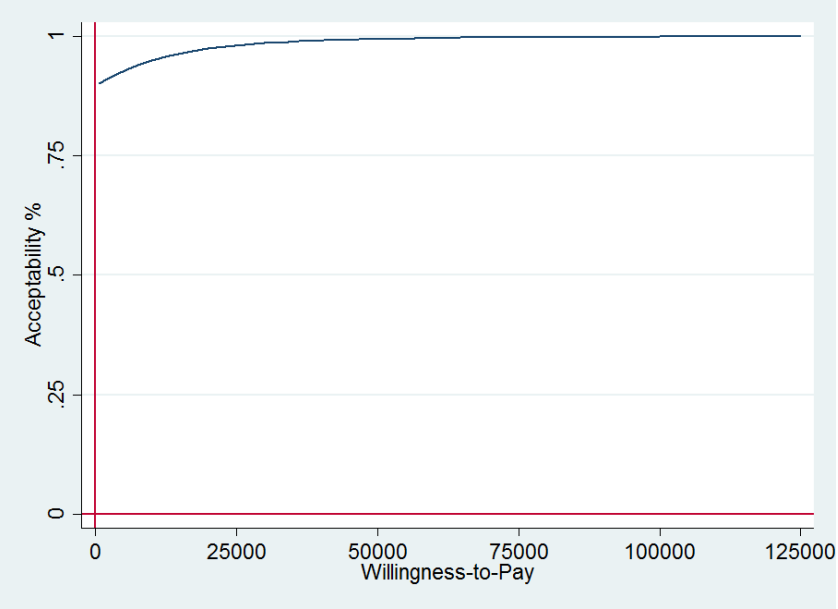


Figure 4A.2 - Cost-Effectiveness Acceptability Curve (Base Case, Period 10)

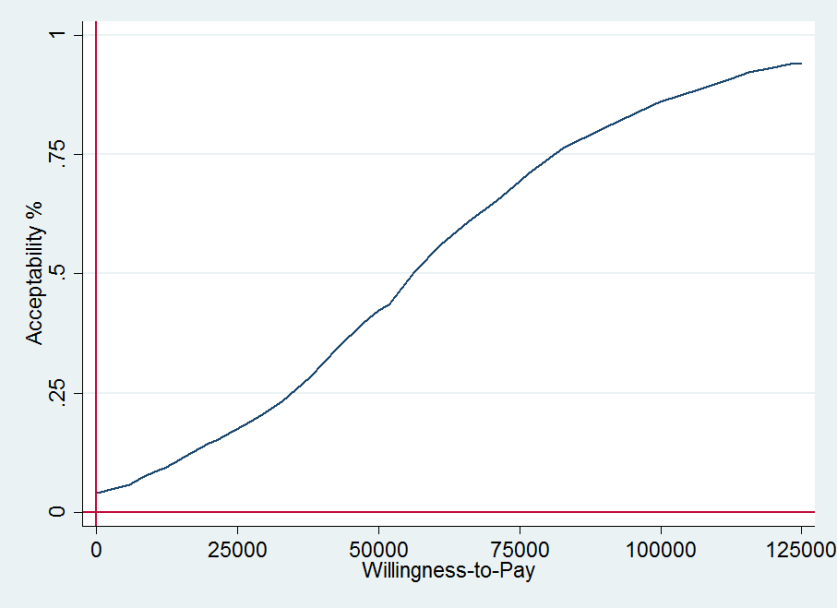


Figure 4A.3 - Cost-Effectiveness Acceptability Curve (Utility -0.05, Period 1)

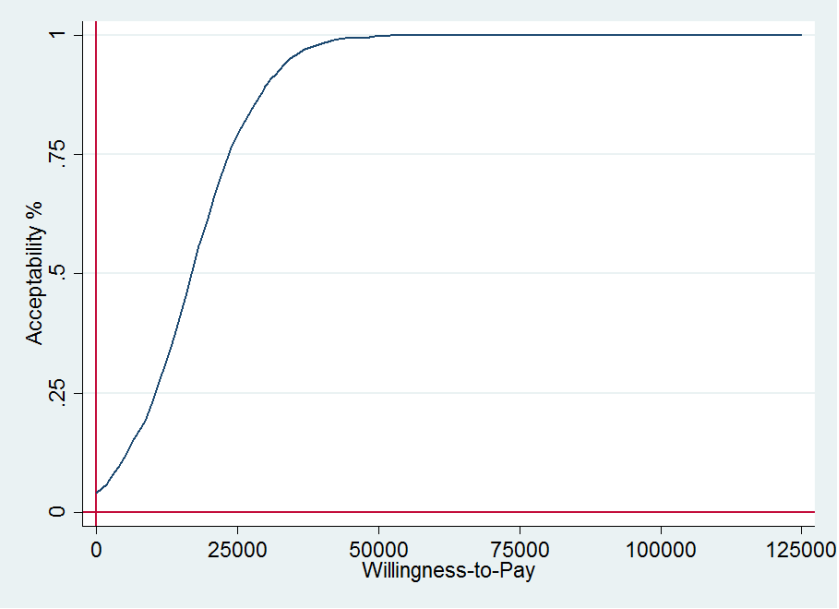


Figure 4A.4 - Cost-Effectiveness Acceptability Curve (Utility -0.16, Period 1)

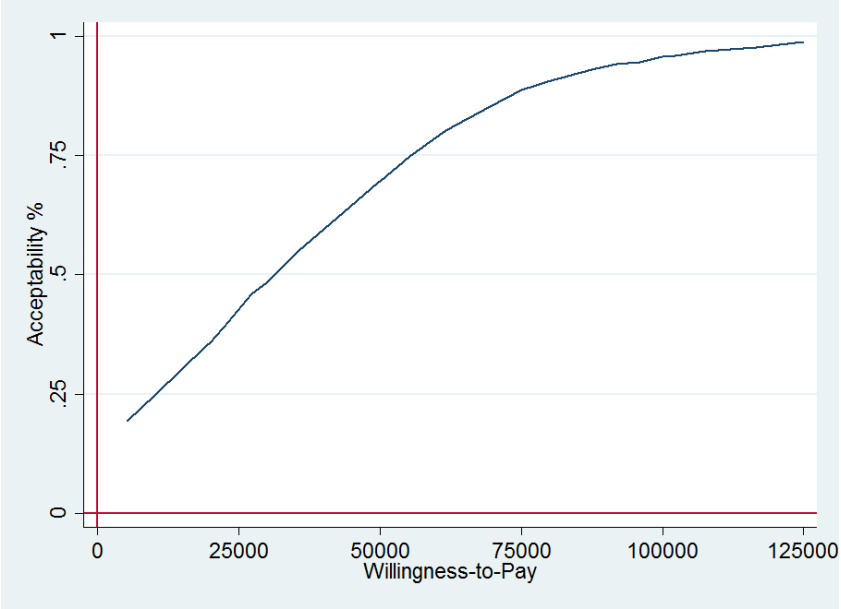


Figure 4A.5 - Cost-Effectiveness Acceptability Curve (Low Intervention Effectiveness, Period 1)

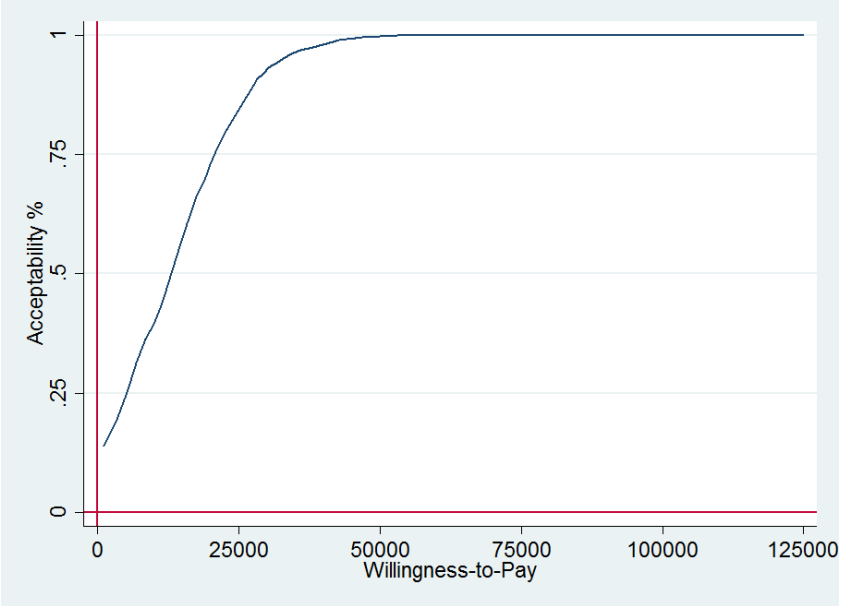


Figure 4A.6 - Cost-Effectiveness Acceptability Curve (High Intervention Effectiveness, Period 1)

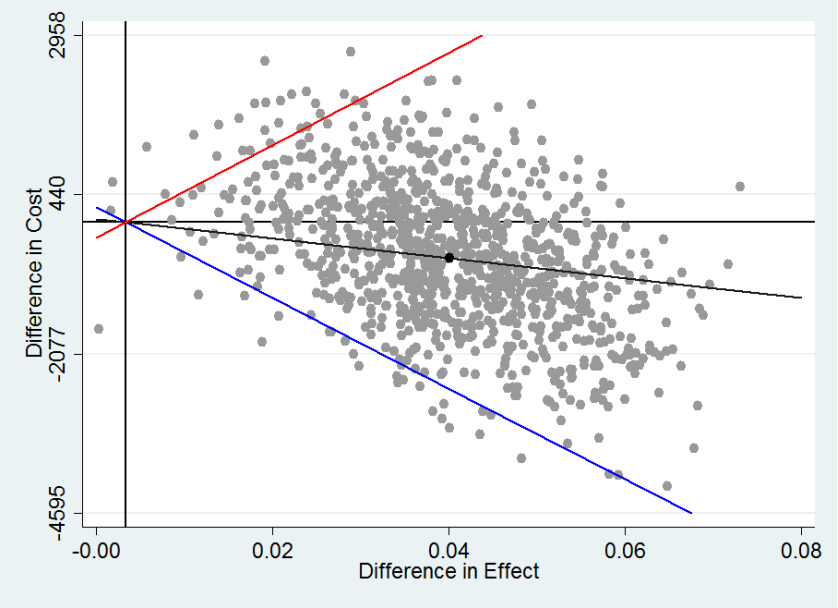


Figure 4A.7 - Bootstrapped ICERs (Low Intervention Effectiveness, Period 5)

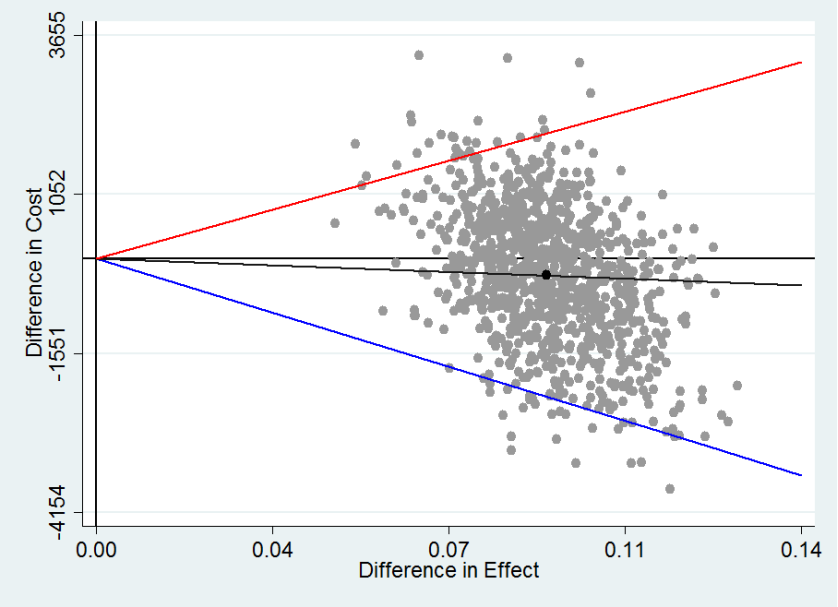


Figure 4A.8 - Bootstrapped ICERs (High Intervention Effectiveness, Period 5)

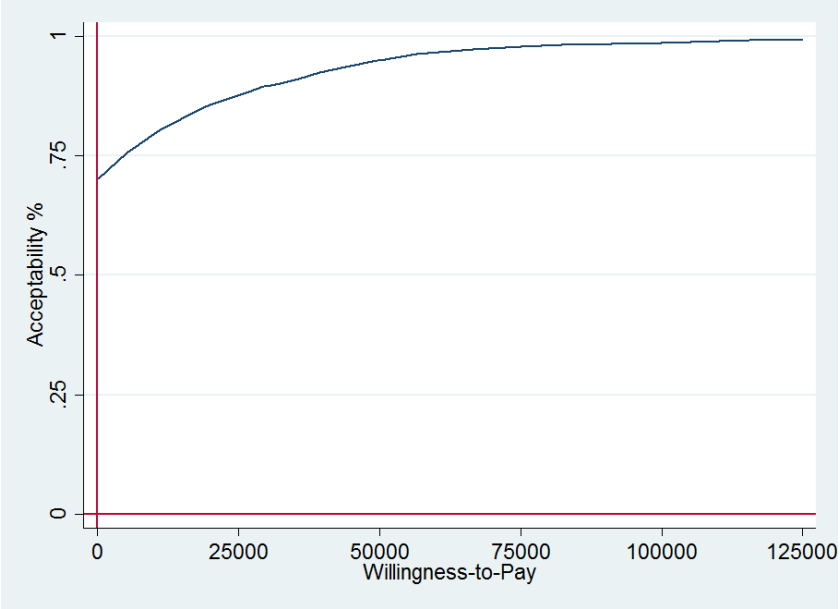


Figure 4A.9 - Cost-Effectiveness Acceptability Curve (Low Intervention Effectiveness, Period 5)

Usual Care Cost | Cost-Effectiveness Acceptability Curve

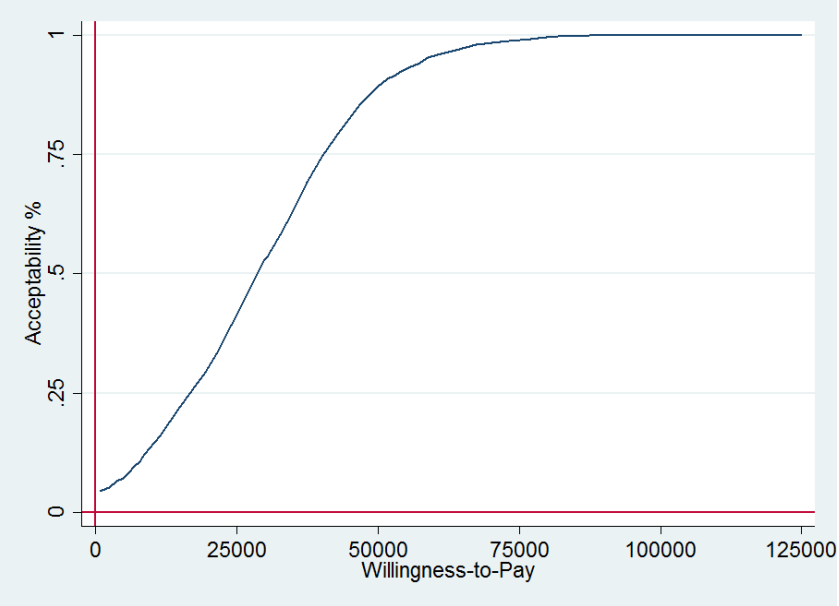


Figure 4A.10 - Cost-Effectiveness Acceptability Curve (Low Usual Care Cost, Period 1)

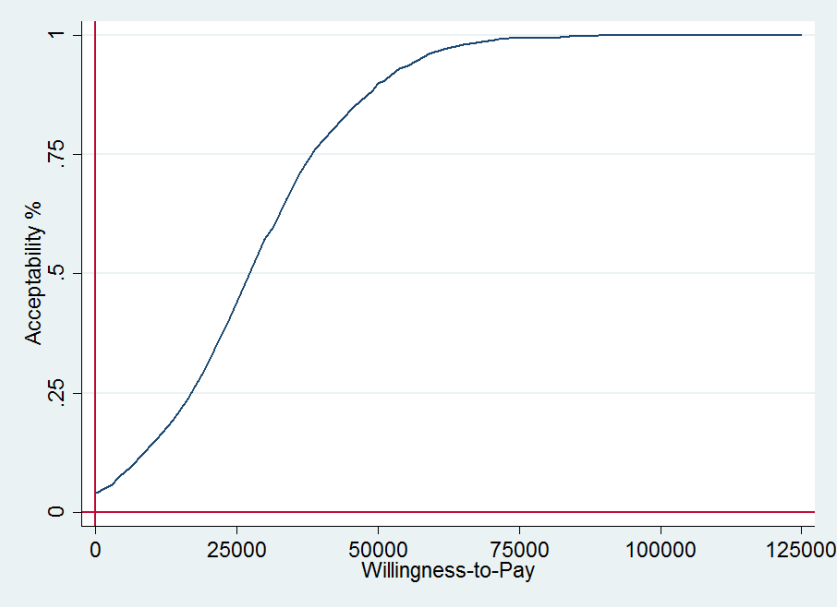


Figure 4A.11 - Cost-Effectiveness Acceptability Curve (High Usual Care Cost, Period 1)

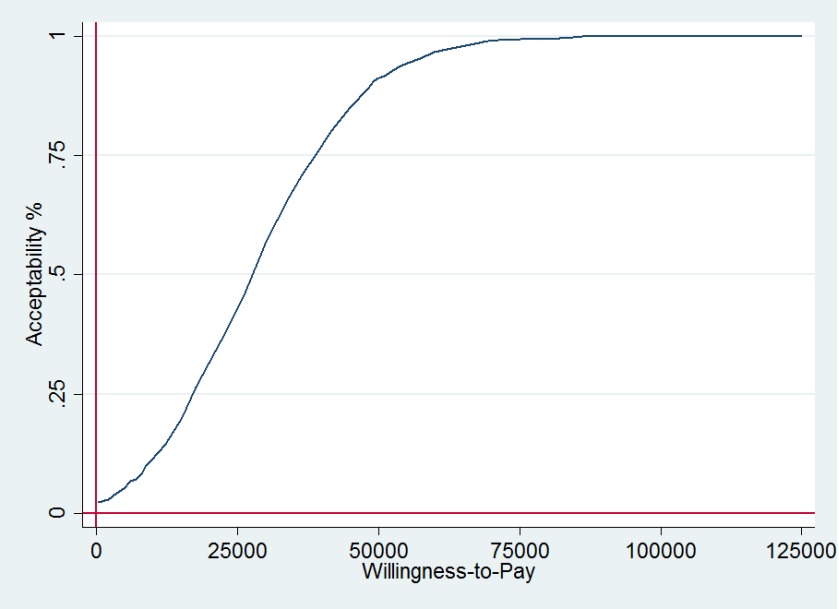


Figure 4A.12 - Cost-Effectiveness Acceptability Curve (Low Usual Care Effectiveness, Period 1)

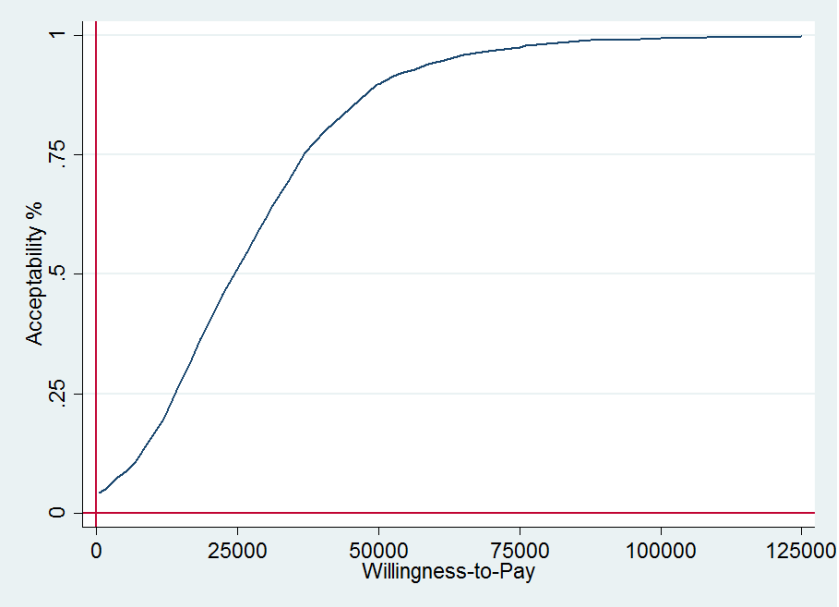


Figure 4A.13 - Cost-Effectiveness Acceptability Curve (Low Usual Care Effectiveness, Period 10)

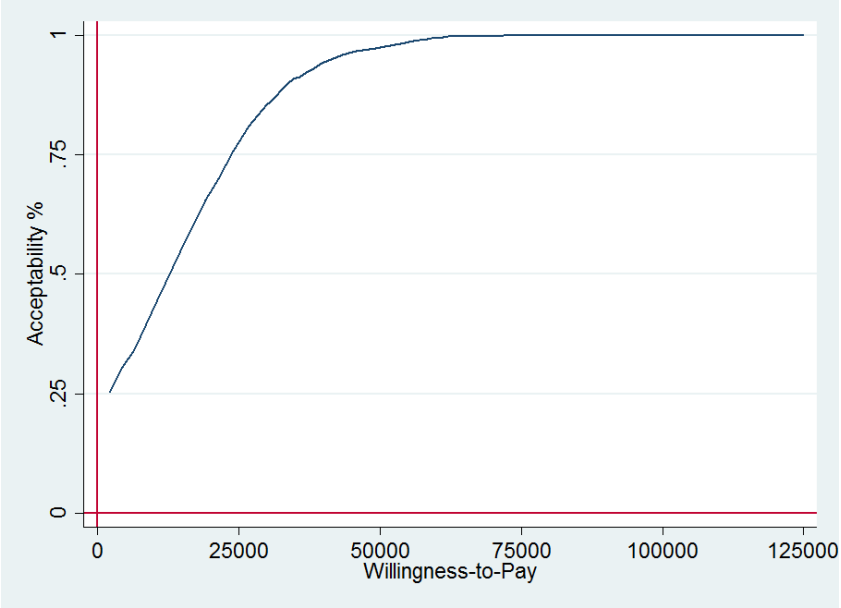


Figure 4A.14 - Cost-Effectiveness Acceptability Curve (High Usual Care Effectiveness, Period 1)

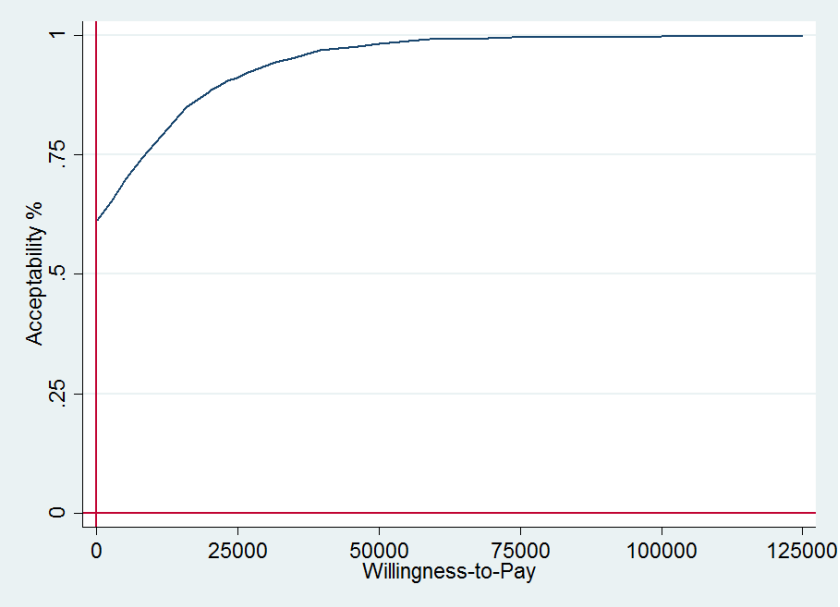


Figure 4A.15 - Cost-Effectiveness Acceptability Curve (High Usual Care Effectiveness, Period 10)

Chapter 5
Conclusion

In this dissertation, we examine two sources of increased health care costs – hospital readmissions and the co-occurrence of depression among individuals with diabetes.

In the first paper, we analyze whether sources of incentive heterogeneity are associated with improvements in hospital performance in the Hospital Readmissions Reduction Program. The results suggest that hospitals that performed poorly in previous years of the program are improving their readmissions significantly more than other hospitals. Further, we develop a novel methodology to calculate the marginal benefit of performance improvements in the program. Comparing actual changes in performance to the calculated marginal benefits, the results suggest that hospitals are decreasing readmissions in most of the conditions where performance improvement reaps the highest benefit in program performance. We also find evidence that there may be a relationship between the share of hospital service volume accounted for by the conditions measured in the program and improvements over time, though this is limited to the conditions that were most recently added to the program. How hospitals interact with incentive programs will continue to grow in importance as reimbursement policies undergo transformations.

The patient population with depression and diabetes is characterized by worse health outcomes, lower health utility, and increased health care costs. In the second paper, we use longitudinal data to develop a model predicting the development of depression among patients

with diabetes. My analysis suggests that based on gender, body-mass index, hypertension, history of stroke, history of heart disease, and duration of diabetes, we can discriminate moderately well between individuals with and without depression. We then build this prediction model into the Michigan Model for Diabetes to improve the clinical accuracy of this diabetes progression simulation model and to allow future research on the potential ways to treat depression within the context of diabetes and the associated complications and comorbidities.

Finally, in the third paper, we use the modifications to the Michigan Model for Diabetes to study the cost-effectiveness of the collaborative care intervention. This treatment approach integrates nurses, physicians, and patients to improve depressive symptomology. Trials have shown that this intervention can be cost-effective in the short-term, but there is no data available regarding the long-term benefits. Using the modified simulation model and a variety of simulation inputs, our analysis finds that the collaborative care intervention can be very good value in both the short- and long-term. Only when the utility loss associated with depression is small or when the effectiveness of the intervention is extremely small does the intervention require a higher willingness-to-pay to be considered cost-effective. Otherwise, in our base-case analysis and a variety of other one-way sensitivity analyses, our model suggests that the collaborative care intervention is very cost-effective when used to treat depression among patients with diabetes.

Future work will continue to evaluate hospital performance in the Hospital Readmissions Reduction Program. As more conditions are added to this program, more hospitals will receive penalties. Subsequently, it could become more difficult for hospitals to avoid penalties, so their responses to the financial incentives may evolve. Further analyses of hospital performance in the new conditions will be informational, as the performance of surgical specialty hospitals may be

hurting the ability of general acute care hospitals to perform well in some conditions. Finally, adding in estimates of the dollar values associated with the penalties hospitals face will provide a more tangible outcome variable to study over time.

In the diabetes and depression realm, future research will work to uncover more details of the development of depression among diabetics. If we can model the length of depression episodes, we should be able to improve the clinical accuracy of our modifications to the MMD. Additionally, many different scenario analyses could be important contributions to the field, including the modeling the administration of the intervention multiple times or adding in a waning function for the usual care treatment. To improve on the memoryless limitation of the model, we will explore incorporating depression history and treatment response history as determinants of future depression experience in the MMD. These changes should help the model better identify individuals with and without depression over time.