Bayesian Hierarchical Difference-in-Differences Models

A DISSERTATION SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL OF THE UNIVERSITY OF MINNESOTA BY

James P. Normington

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

Advised by Dr. Eric F. Lock

Co-advised by Dr. Thomas A. Murray

December 2019

Acknowledgements

I am deeply grateful to my adviser, Eric Lock, for celebrating my triumphs while constructively criticizing my missteps throughout the completion of this work. His well-known excellence in teaching is matched by his advising, and my understanding of statistics has vastly improved by working with him. I would like to also acknowledge Thomas Murray for his role as co-adviser. His insights always caused me to cast a more critical eye at the approaches I took and the claims I made. I would also like to thank Kevin Peterson and the entire UNITED team for providing the data and research questions which both inspired much of the work in this dissertation and partially funded my studies at the University. Special thanks go to Caroline Carlin, who both served on my committee and offered excellent feedback (both with respect to our applied problems and statistical methodology) throughout these years. I would like to thank David Vock for serving on my committee; I am particularly thankful for his excellent feedback during my preliminary oral examination which I feel has greatly improved this work. Finally, I will forever be indebted to the University of Minnesota, the School of Public Health, and the Division of Biostatistics; thank you for believing in me.

Dedication

To my parents Carolyn Ann Koleske and Sydney James Normington, for selflessly providing opportunities which have enriched my life more than I could ever repay.

To my sister and childhood best friend Julia Kathryn Schwartz, for showing me that a fiery resolve defeats all demons, no matter how great.

To my wife Angela Christine Normington, for being an unconditional and infinite source of support, encouragement, and love during graduate school. From our first meeting through marriage and beyond, my favorite place will always be in the passenger seat as you are driving me home.

Abstract

A popular method for estimating a causal treatment effect with observational data is the difference-in-differences (DiD) model. In this work, we extend the classical DiD model to the hierarchical context in which data cannot be matched at the most granular level. Our motivating example is an application to assess the impact of primary care redesign policy on diabetes outcomes in Minnesota, in which patient-level outcomes are not matched longitudinally, and thus the mean change in outcome, treatment(s), and covariates are measured at the clinic level. We propose a Bayesian *hierarchical difference-in-differences* (HDiD) model which estimates the treatment effect by regressing the treatment on a latent variable representing the mean change in group-level outcome. We first apply the HDiD model to measure the impact of primary care redesign on clinics in Minnesota from 2008 to 2012. We go on to present theoretical and empirical results showing that an HDiD model that fails to adjust for a particular class of confounding variables biases the treatment effect estimate. Motivated by the need for covariate adjustment, we propose and implement various approaches to perform variable selection using a structured Bayesian spike-and-slab model in the HDiD context, evaluating their properties with theoretical results and through simulation. We then conduct a cross-sectional analysis to measure the specific primary care services and resources which most strongly relate with improved diabetes outcomes in 2017. We conclude by introducing an additional variable selection approach which leverages the temporality of the HDiD context, as well as any structure in the predictors.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

A central goal of biostatistics is to measure the impact of a treatment on a health outcome. The degree to which a treatment directly affects (rather than merely correlates with) an outcome is a causal question, in which randomized controlled trials are the gold standard. However, due to impracticality and potential ethical issues, randomized controlled trials are rarely used to evaluate large-scale policies.[1] Thus, we are often confined to using observational studies to make causal inferences.

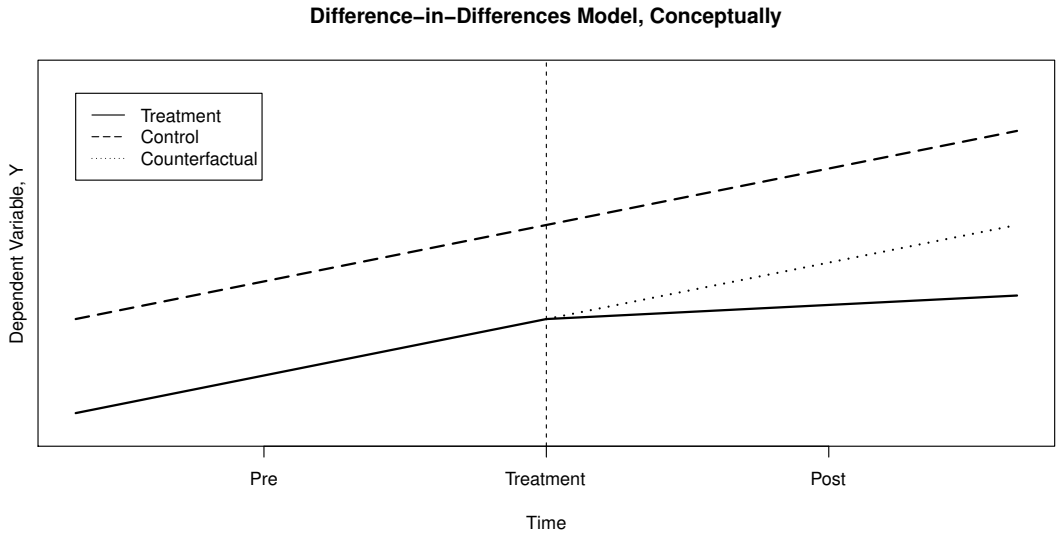


Figure 1.1: A difference-in-differences model, conceptually; lines signify the observed values in the treatment group, the observed values in the control group, and the values the treatment group would have would have observed had they not received the treatment (counterfactual)

A major issue in making causal inferences from observational studies is the virtually unavoidable presence of confounding variables. A popular observational method that avoids the effect of static confounders, or confounding variables whose values and relation to the outcome do not change over time, is the difference-in-differences (DiD) model. DiD estimation first defines a treatment and an outcome. Under the classical setting the treatment is binary, defining two groups (e.g., "treatment" and "control"). The standard DiD model tests for a difference between the average change in outcome over time in the treatment group and the average change in outcome over time in the control group. Specifically, it takes the difference in mean outcome between the groups *before* the treatment (Difference 1), takes the difference in mean outcome between the groups *after* the treatment (Difference 2), and then takes the difference between those two differences (Difference 2 - Difference 1).

The DiD model is used often in econometrics, social science, and marketing. As a canonical example of applying the DiD model, Card & Krueger compared the change in employment in New Jersey vs. Pennsylvania after New Jersey adopted an increase in the minimum wage; in this context New Jersey could be considered the treatment group and Pennsylvania the control group.[2] Figure 1.1 illustrates an example of a DiD model.

Formally, let $Y_i^{(t)}$ denote the outcome variable for subject $i = 1, ..., n$ at timepoint $t \in \{0, 1\}$, where $t = 0$ indicates the measurement was taken before the treatment and $t = 1$ indicates the measurement was taken after the treatment. Let $T_i^{(t)}$ denote the treatment status for individual i at timepoint t. The observed treatment level for individual i is then $T_i = T_i^{(1)} - T_i^{(0)}$. With a binary treatment, the "treatment" group's treatment status is $T_i = T_i^{(1)} - T_i^{(0)} = 1 - 0 = 1$ and the "control" group's treatment status is $T_i = T_i^{(1)} - T_i^{(0)} = 0 - 0 = 0$. In the continuous case, subject i's treatment status is $T_i = T_i^{(1)} - T_i^{(0)}$. A common DiD model arises as

$$Y_i^{(t)} = \beta_0(1 - T_i) + \beta_0' T_i + \phi \mathbb{1}(t = 1) + \Delta T_i \mathbb{1}(t = 1) + \epsilon_i^{(t)} \tag{1.1}$$

where β_0 and β_0' are the pre-treatment means of the control and treatment groups respectively, ϕ

is the common time trend assumed for every subject, $\mathbb{1}$ is the indicator function, Δ is the treatment effect of interest, and $\epsilon_i^{(t)}$ is the subject's normally distributed residual at timepoint t. In this formulation, Δ is the average treatment effect in the post-treatment timepoint.

In addition to the standard Gauss-Markov assumptions for an ordinary linear model, the model in (1.1) explicitly makes an assumption of *parallel trends*.[3] The parallel trends assumption states that, absent treatment, the outcomes of the treatment and control groups are expected to change at the same rate; in (1.1), the assumed common rate is ϕ. The parallel "Treatment" and "Control" lines pre-treatment and parallel "Control" and "Counterfactual" lines post-treatment in Figure 1.1 illustrate the parallel trends assumption.

The typical DiD estimator is the observed difference between the average observed trend in the treated group and the average observed trend in the control group: $\hat{\Delta} = \left(\bar{Y}^{(1)}_{T.=1} - \bar{Y}^{(0)}_{T.=1}\right) - \left(\bar{Y}^{(1)}_{T.=0} - \bar{Y}^{(0)}_{T.=0}\right)$. If certain identifying assumptions are satisfied, then $\hat{\Delta}$ is an unbiased and consistent estimator of Δ.[4] Using $\hat{\Delta}$ to estimate Δ differences out the effect of static confounders, eliminating the need to adjust the model for them. However, an important identifying assumption is the absence of *dynamic confounders*, which are confounding variables whose values or relation to the outcome change over time.

For model (1.1), it is often sufficient to consider the change in outcome for each individual:

$$Y_i^{\text{diff}} = \phi + \Delta T_i + \epsilon_i \tag{1.2}$$

where $Y_i^{\text{diff}} = Y_i^{(1)} - Y_i^{(0)}$ and $\epsilon_i = \epsilon_i^{(1)} - \epsilon_i^{(0)}$. This is a classical linear model, in which estimation and inference proceeds with well-understood theoretical results. In this work, we instead consider the hierarchical context in which change is not observed at the most granular level (e.g., patient-level outcomes are not observed, but a patient-level treatment effect is desired). Further, we generalize (1.1) to allow for multiple (potentially continuous) treatments. Our motivating example is an application to assess the impact of primary care redesign policy on diabetes outcomes at clinics in Minnesota, in which individual outcomes are not matched longitudinally. Thus, the change in

outcome, treatment(s), and covariates are measured at the clinic level.

In what follows, we propose a Bayesian hierarchical version of the DiD model and discuss the causal estimand and necessary assumptions for identification. With this new model as our centerpiece, we go on to apply it to a motivating data set, explore the consequences of mis-specifying this model, characterize which type of covariates this model needs to be adjusted for, and propose variable selection algorithms in this model's framework. Chapter 2 introduces the research question and data sets that motivated the construction of the Bayesian hierarchical DiD model. Chapter 3 states the Bayesian hierarchical DiD model formally and, with a few context-specific modifications, applies it to the data introduced in Chapter 2. Chapter 4 presents empirical and theoretical biases incurred by model misspecification, explores covariate adjustment and proposes methods of performing variable selection in this context, applies the methods to a more recent version of the data introduced in Chapter 2, and investigates how estimation of the causal effect behaves asymptotically under different model choices and variable selection techniques. Chapter 5 performs a cross-sectional analysis of our motivating data at a more granular level than Chapters 3 and 4. Chapter 6 proposes an additional variable selection approach which leverages the temporality and data structure in this context. Chapter 7 concludes the work. Notation and sampling algorithms are relegated to appendices.

Chapter 2

Motivating Data

Diabetes mellitus type 2, is a chronic condition that impacts the way the body breaks down glucose, its main source of energy. Diabetes can negatively impact one's endocrine, excretory, digestive, kidney, circulatory, integumentary, central nervous, and reproductive systems.[5] A person with type 2 diabetes either fails to respond normally to or has low levels of insulin, the hormone most responsible for regulating the amount of glucose in the blood. Some people with type 2 diabetes can manage their blood sugar levels with diet, exercise, sleeping habits, and smoking habits, while many also require insulin administration to maintain their blood sugar levels.[6] The glycated hemoglobin (A1c) test is the standard measure of one's average blood sugar for the previous two to three months, and is very indicative of how a patient is managing her diabetes. To identify trends in a type 2 diabetes patient's blood sugar levels, the American Diabetes Association recommends routine A1c tests and consultations with their primary care physician.[7]

Primary care practices provide a structured system for individuals to manage their diabetes.[8,9] However, national quality of care metrics show that significant gaps exist between evidence-based optimal diabetes care and the diabetes care actually delivered to patients.[10] To increase the effectiveness of primary care delivery, new models that increase coordination, emphasize prevention, and enhance collaboration between multidisciplinary teams have been proposed.[11] In 2008, the Minnesota State Legislature endorsed the Patient Centered Medical Home (PCMH) as the preferred model for

primary care redesign.[12] For a clinic to retain state certification as a PCMH, it must demonstrate maintaining a comprehensive set of resources and services for patients validated through onsite state inspection, annual reporting, and regular tracking. These services and resources intend to satisfy the United States Department of Health and Human Services' definition of the PCMH, which encompasses five pillars of primary care: comprehensive care (the PCMH addresses a patient's full health profile), patient-centered care (respecting the patient's personal needs, values, and preferences), coordinated care (connecting specialty care, hospital care, home care, etc.), accessible services (online scheduling, 24/7 telephone and electronic access to care team), and quality and safety (commitment to improving quality with evidence-based care).[13] This legislated process provided the setting for a natural experiment to compare the delivery of standardized and validated clinical services on diabetes outcomes over the last ten years.

A group of researchers from Medica Research Institute, HealthPartners Institute, and the University of Minnesota's Department of Family Medicine and Division of Biostatistics, referred to hereafter as UNITED, was organized to investigate the effect of clinical services and resources associated with PCMH certification on diabetes outcomes. To do so, we leverage a decade's worth of outcome data from MN Community Measurement (described in Section 2.1), neighborhood-level socioeconomic data from the American Community Survey (Section 2.2), and clinic resources and services survey data from the Physician Practice Connections-Research Survey (Section 2.3).

2.1 MN Community Measurement Optimal Diabetes Care

As a result of the 2008 legislation, the Minnesota Department of Health contracted with the nonprofit organization MN Community Measurement (MNCM) to measure quality of care for patients living with diabetes. Prior to the legislation, MNCM had been collecting diabetes quality of care measures from clinics participating on a voluntary basis; the change in legislation mandated diabetes reporting for all primary care clinics providing care to 30 or more patients with diabetes.

The MNCM Optimal Diabetes Care dataset contains the primary measures of optimal diabetes

care incorporated into the MNCM "D5" standard of care for patients living with diabetes. The five criteria include low-density lipoprotein (LDL) below 100 mg/dL, systolic blood pressure (SBP) below 140/90 mmHg, hemoglobin A1c below 8%, daily use of aspirin or other antiplatelet medication if diagnosed with ischemic vascular disease (IVD), and documentation of non-smoking status. The data have a three-level hierarchical structure, as patient-year data are matched to the primary clinic where most measurements were taken that year, and each clinic belongs to a broader collection of clinics we refer to as a "system". MNCM contains data for years 2008-2017, but our dataset does not match patients across years so a standard longitudinal analysis is not feasible.

2.2 American Community Survey

The American Community Survey (ACS) is an annual survey conducted by the U.S. Census Bureau which administers a questionnaire to a sample of addresses capturing many of the variables included in the long form decennial census. We use the survey results aggregated to the ZIP code summary level, matched to patient ZIP code to describe the environment in which the patient lives and functions. Using confirmatory factor analysis, Swaney[14] computes measures of socioeconomic status in the patients' neighborhoods (income and wealth) as a function of education, housing costs, use of the Supplemental Nutritional Assistance Program (SNAP), household income, and family structure. We also capture the influence of race and ethnicity by including the percentage of the ZIP code residents who identify as white, non-Hispanic.

2.3 Physician Practice Connections-Research Survey

The Physician Practice Connections-Research Survey[15] is a survey designed to measure primary care organizational infrastructure across five of the six domains of Bodenheimer and Wagner's Chronic Care Model (CCM): health care organization, delivery system redesign, clinical information systems, decision support, and self-management support.[16] Each question on the PPCRS in turn corresponds to one of these five CCM domains. The PPCRS was sent to all 111 Minnesota

primary care sites that had been certified as health care homes (Minnesota's version of a PCMH) in 2011. Respondents were asked to report organizational structure at present (2011) and (by recall) in 2008. The PPCRS was again sent out in 2017 to all Minnesota primary care practices complying with 2016 MNCM reporting, whether or not they were certified as a PCMH.

2.4 Modeling Considerations

In our application, we are primarily interested in modeling the effect of clinic-level redesign on diabetes management outcomes. To achieve this, we use the patient-year MNCM data for our outcomes, the clinic-year PPCRS results as our main exposure of interest, and we adjust for potential confounders using the MNCM and ACS data. As we are interested in how the *change* in primary care redesign relates with the *change* in diabetes outcomes, a DiD model with a continuous treatment is a natural approach to pursue. However, these data have several characteristics that must be carefully considered from a modeling perspective, including the following:

1. *Inability to match patients*: Patient data were not matched across years, so any model must measure the relationship between maturity in primary care transformation and mean change in diabetes outcomes, averaged across patients at the *clinic* level. However, the variability of the mean difference in clinic outcomes and covariates can differ depending on the number of patients observed at a clinic and the heterogeneity of the patient population within a clinic; this clinic-level heteroscedasticity must be accounted for.

2. *Hierarchical structure*: The outcome and predictors are hierarchical, as patients visit clinics which exist within health care systems. Clinics that belong to the same system organization likely have similar policies and electronic health record systems, and may also be similar geographically, economically, and demographically. It is unreasonable to assume mean diabetes outcomes across clinics within a system are uncorrelated, so applying a Bayesian hierarchical model is a natural way to account for this structure.

3. *Confounding*: While the typical DiD approach inherently balances static confounders, the role of confounding in the hierarchical difference-in-differences context is less clear.

Chapter 3

Bayesian Hierarchical Difference-in-Differences Models

In this chapter, we describe a comprehensive Bayesian hierarchical difference-in-differences model. We introduce the causal estimand of interest and necessary assumptions for its identification. We then go on to provide a step-by-step description of how our model evolved as we considered different facets of the data and analytic goals motivated by the research question in Chapter 2.

3.1 A Bayesian Hierarchical Difference-in-Differences Model

With the modeling considerations at the end of Section 2.4 in mind, one could extend the standard DiD model in Equation (1.1) to allow for multiple (potentially continuous) treatment levels (e.g., dose-response relationships) as well as adjust for group-level dynamic confounders in a hierarchical

fashion:

$$
\begin{aligned}
Y_{ji}^{(0)} &\sim N(\mu_j, \widetilde{\sigma}_j^2) \\
Y_{ji}^{(1)} &\sim N(\mu_j + \mu_j^{\text{diff}}, \sigma_j^2) \\
\mu_j &\sim N(T_j\widetilde{\Delta} + \boldsymbol{X}_j^T\widetilde{\boldsymbol{\beta}}, \widetilde{\tau}^2) \\
\mu_j^{\text{diff}} &\sim N(T_j\Delta + \boldsymbol{X}_j^T\boldsymbol{\beta}, \tau^2),
\end{aligned}
\tag{3.1}
$$

where j indexes groups (e.g., clinics), $j = 1, ..., J$, μ_j is the pre-treatment mean outcome of group j, μ_j^{diff} is the mean change in outcome of group j from $t = 0$ to $t = 1$, and $\boldsymbol{X}$ is a design matrix comprised of group-level covariates. $\widetilde{\sigma}_j^2$ and σ_j^2 are group-j specific outcome variances pre- and post-treatment, respectively, T_j is the treatment exposure of group j, and $\widetilde{\tau}^2$ and τ^2 are the variances of the pre- and post-treatment mean outcomes, respectively. To measure multiple treatment effects, one could alter Equation (3.1) extending the vector of observed treatment levels $\boldsymbol{T}$ to a matrix of group-level exposures and extending the single treatment effect Δ to a vector of treatment effects $\boldsymbol{\Delta}$. We focus on estimating one treatment effect throughout this chapter and Chapter 4.

Hereafter, we refer to Equation (3.1) as a *hierarchical difference-in-differences model* (HDiD), and the third and fourth lines of Equation (3.1) as the "baseline" and "change" models, respectively. An analogous model was first introduced in Normington et al.,[17] with a discussion of required assumptions for this model to identify the causal effect. Equivalent multivariate representation of the baseline and change models are

$$
\begin{aligned}
\boldsymbol{\mu} &\sim \boldsymbol{N_J}(\boldsymbol{T}\widetilde{\Delta} + \boldsymbol{X}\widetilde{\boldsymbol{\beta}}, \widetilde{\boldsymbol{\Omega}}) \\
\boldsymbol{\mu}^{\text{diff}} &\sim \boldsymbol{N_J}(\boldsymbol{T}\Delta + \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{\Omega}),
\end{aligned}
$$

where for now we set $\widetilde{\boldsymbol{\Omega}} = \widetilde{\tau}^2\boldsymbol{I_J}$ and $\boldsymbol{\Omega} = \tau^2\boldsymbol{I_J}$.

After specifying the likelihood in (3.1), we choose to proceed with Bayesian inference for a few

practical and philosophical reasons. While maximizing this likelihood may be possible, the heteroscedastic clinic-level outcome variances $\widetilde{\sigma}_j^2$ and σ_j^2 make this estimation challenging. Furthermore, the hierarchical structure of the MNCM and PPCRS data naturally lends itself to a Bayesian hierarchical model. After assigning priors, we can compute the posterior distribution of each parameter with standard Markov chain Monte Carlo (MCMC) techniques (e.g., conditionally conjugate Gibbs sampling, with careful prior choices). Bayesian inference also allows us to incorporate prior knowledge, if available, about the model parameters, most importantly Δ in this context.

Note that in (3.1), we model pre-treatment means $\boldsymbol{\mu}$ with the same predictors $\boldsymbol{X}$ used to model the change in pre- to post- treatment means $\boldsymbol{\mu}^{\textbf{diff}}$. This may be counter-intuitive, especially since some covariate values given in $\boldsymbol{X}$ could be realized after the pre-treatment timepoint. However, without variable selection, it is critically important that $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$ are modeled with the same set of predictors; otherwise, estimation of Δ may be biased if we adjust $\boldsymbol{\mu}^{\textbf{diff}}$, but not $\boldsymbol{\mu}$, for variables predictive of $\boldsymbol{\mu}$. We expound on this phenomenon in Chapter 4.

3.2 Causal Estimand and Identification

In (3.1), our primary parameter of interest Δ resolves to $\Delta = \mathbb{E}\Big[Y_{ji}^{(1)} - Y_{ji}^{(0)} \mid T_j = a+1, \boldsymbol{X}_j^T = \boldsymbol{x}_j\Big] - \mathbb{E}\Big[Y_{ji}^{(1)} - Y_{ji}^{(0)} \mid T_j = a, \boldsymbol{X}_j^T = \boldsymbol{x}_j\Big]$. To define Δ as a causal estimand, we use the potential outcomes notation introduced in the Rubin causal model.[18] Letting $Y_{ji}(T_j = a)$ denote the potential outcome of subject i had group j received treatment level a, we define Δ as $\Delta = \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = 1) - Y_{ji}^{(1)}(T_j = 0)\Big]$ in the binary treatment case. More generally, the treatments T_j may be non-binary or continuous, in which Δ gives the expected difference in post-treatment outcome for an individual had they received an additional unit of exposure:

$$\Delta = \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(1)}(T_j = a)\Big]. \tag{3.2}$$

In either setting, only one of the potential outcomes $\left\{ Y_{ji}^{(1)}(T_j = a) \mid a \in \mathcal{T} \right\}$, where $\mathcal{T}$ denotes the support of $\boldsymbol{T}$, is observed. The first assumption required for (3.1) to identify (3.2) is *Exogeneity*, which states that realized values of the covariates are not changed by the treatment exposure:[19]

Assumption 1 (*Exogeneity*)**.** $\boldsymbol{X}(\boldsymbol{T} = \boldsymbol{a}) = \boldsymbol{X}(\boldsymbol{T} = \boldsymbol{a'})$ *for any two values* $\boldsymbol{a}, \boldsymbol{a'}$ *of* $\boldsymbol{T}$.

Just as in the typical DiD setup, the model in 3.1 also requires an assumption of parallel trends. Although in our motivating context, the $Y_{ji}^{(1)}(T_j = a) - Y_{ji}^{(0)}(T_j = a)$ terms are unobservable (recall we are unable to match patients across time), we must assume their potential outcomes are equal across observed T_j in expectation. We thus extend the parallel trends assumption,[4,20,21] which is typically concerned with the potential outcomes of two groups defined by a binary exposure, to our context with no proper control group and a continuous exposure:

Assumption 2 (*Parallel Trends*)**.** $\mathbb{E}\left[Y_{ji}^{(1)}(T_j = a) - Y_{ji}^{(0)}(T_j = a) \mid T_j = a, \boldsymbol{X}_j^T = \boldsymbol{x} \right] = \mathbb{E}\left[Y_{ji}^{(1)}(T_j = a) - Y_{ji}^{(0)}(T_j = a) \mid T_j = a', \boldsymbol{X}_j^T = \boldsymbol{x} \right]$, *for any two values* a, a' *of* T_j.

The final assumption required to identify (3.2) is the *No Anticipatory Behavior* assumption,[4,20] which states that the future treatment level has no effect on any two pre-treatment potential outcomes:

Assumption 3 (*No Anticipatory Behavior*)**.** $\mathbb{E}\left[Y_{ji}^{(0)}(T_j = a) - Y_{ji}^{(0)}(T_j = a') \mid \boldsymbol{X}_j^T = \boldsymbol{x} \right] = 0$, *for any two values* a, a' *of* T_j.

Now, we present the main statement regarding identifiability.

Theorem 1. *If Assumptions 1-3 hold, then the* Δ *as specified by Model* (3.1) *identifies the estimand* $\mathbb{E}\left[Y_{ji}^{(1)}(T_j = a + 1) - Y_{ji}^{(1)}(T_j = a) \right]$.

Proof. We show the result directly:

$$\Delta = \mathbb{E}\left[Y_{ji}^{(1)} - Y_{ji}^{(0)} \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a + 1 \right] - \mathbb{E}\left[Y_{ji}^{(1)} - Y_{ji}^{(0)} \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a \right]$$
$$= \mathbb{E}\left[Y_{ji}^{(1)} - Y_{ji}^{(0)} \mid \boldsymbol{X}_j^T = \boldsymbol{x}(a+1), T_j = a + 1 \right]$$

$$
\begin{aligned}
&\quad - \mathbb{E}\Big[Y_{ji}^{(1)} - Y_{ji}^{(0)} \mid \boldsymbol{X}_j^T = \boldsymbol{x}(a), T_j = a\Big] \text{ by Assumption 1} \\
&= \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(0)}(T_j = a+1) \mid \boldsymbol{X}_j^T = \boldsymbol{x}(a+1), T_j = a+1\Big] \\
&\quad - \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a) - Y_{ji}^{(0)}(T_j = a) \mid \boldsymbol{X}_j^T = \boldsymbol{x}(a), T_j = a\Big] \\
&= \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(0)}(T_j = a+1) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a+1\Big] \\
&\quad - \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a) - Y_{ji}^{(0)}(T_j = a) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a\Big] \text{ by Assumption 1} \\
&= \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(0)}(T_j = a+1) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a+1\Big] \\
&\quad - \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a) - Y_{ji}^{(0)}(T_j = a) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a\Big] \\
&\quad + \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(0)}(T_j = a+1) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a\Big] \\
&\quad - \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(0)}(T_j = a+1) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a\Big] \\
&= \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(0)}(T_j = a+1) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a\Big] \\
&\quad - \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a) - Y_{ji}^{(0)}(T_j = a) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a\Big] \text{ by Assumption 2} \\
&= \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(1)}(T_j = a) + Y_{ji}^{(0)}(T_j = a) - Y_{ji}^{(0)}(T_j = a+1) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a\Big] \\
&= \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(1)}(T_j = a) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a\Big] \text{ by Assumption 3}
\end{aligned}
$$

So, under the assumptions of our statistical model and Assumptions 1-3, $\Delta = \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(1)}(T_j = a) | \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a\Big]$ $\forall\{\boldsymbol{x}, a\}$. Finally,

$$
\begin{aligned}
\Delta &= \mathbb{E}_{\boldsymbol{x}}[\Delta] \\
&= \mathbb{E}\Big\{\mathbb{E}_{\boldsymbol{x}}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(1)}(T_j = a) \mid \boldsymbol{X}_j^T = \boldsymbol{x}, T_j = a\Big]\Big\} \\
&= \mathbb{E}\Big[Y_{ji}^{(1)}(T_j = a+1) - Y_{ji}^{(1)}(T_j = a)\Big]. \blacksquare
\end{aligned}
$$

3.3 Application to UNITED Data

We return to the research question proposed in Chapter 2; that is, we intend to measure how primary care redesign over time relates with changes in diabetes outcomes in the state of Minnesota from 2008 to 2011. The notation in this section is particularly dense, so we refer the reader to Appendix A, in which Tables A.1 and A.2 describe the notation entirely.

Factors that influence diabetes outcomes are complex and likely interrelate. In addition, factors influencing a clinic's decision about which services and resources to implement may correlate with diabetes outcomes, making causal inference challenging. A cross-sectional analysis approach is therefore limited, even with propensity adjustment[22] or other techniques, because of the potential for unmeasured confounding. Instead, we use data from before and after PCMH certification to assess the association between changes in a composite measure of clinic services and changes in diabetes outcomes, to account for confounding from shifting clinic characteristics. We apply the model proposed in (3.1) separately to three diabetes management outcomes $\boldsymbol{Y}$: patients' last recorded A1c test scores (A1c), low-density lipoprotein levels (LDL), and systolic blood pressure (SBP) for the year. The clinic-level candidate predictors $\boldsymbol{X}$ include a continuous metric of clinic transformation maturity based on PPCRS responses that will be explained more thoroughly in Section 3.3.1, as well as patient-year-level covariates insurance type indicator (1 if commercial insurance, 0 otherwise), age, biological sex, and indicators of whether or not the patient was diagnosed with ischemic vascular disease (IVD), each of which are averaged from the patient level up to the clinic level. Recall that we have PPCRS data available in 2011 and (by recall) 2008, so we use these years to measure clinic transformation maturity from pre- to post-treatment. However, it may be unreasonable to assume patients were affected by primary care restructure in the same year the Minnesota State Legislature endorsed the PCMH model, so we lag both pre- and post- treatment PPCRS predictors by one year to allow for any causal effect from the policy to manifest. Thus, we focus on patient outcomes in 2009 as the pre-treatment measurements and 2012 as the post-treatment measurements, denoted by $\boldsymbol{Y}^{(2009)}$ and $\boldsymbol{Y}^{(2012)}$ respectively.

Patient-year records containing implausible values for these outcomes were omitted from the analysis. Specifically, we only average patient-years with A1c scores between 3 and 25 percent, LDL values between 0 and 1000 mg/dL, SBP values greater than 50 mmHg and greater than diastolic blood pressure (DBP), where DBP is greater than 0 mmHg, excluding just under 5% of total patient-years. We were interested in measuring improvements in type 2 diabetes management, so patients identified as having type 1 diabetes were also excluded from the analysis.

3.3.1 Quantifying Clinic Transformation Maturity

In the most typical causal analysis, a group's treatment status is binary; that is, Δ is the effect of receiving the treatment relative to not receiving the treatment. In the context outlined in Chapter 2, this is not the case. Only clinics certified by the state of Minnesota as a health care home within the first year of the health care home legislation's effective date were sent the PPCRS survey for completion, so its population represents a pool of "early adopter" clinics. In this way, there is no proper "treatment" and "control" groups, so the PPCRS responses are used to infer a continuous treatment status. This inferred treatment status represents how mature the clinic is in the transformation of its primary care delivery. Mathematically, we define clinic j's "clinic score" (c_j) as the score from the first principal component,[23] i.e., the corresponding element of the first left-singular vector from the singular value decomposition of the row-centered PPCRS response matrix, where items with options "No", "Yes, needs improvement", "Yes, works well", and "Don't know" are coded as $\{1, 2, 3, \text{NA}\}$, respectively, and items with options "No" and "Yes" coded as 0 and 1, respectively. Since we are interested in modeling how clinic j's maturity in primary care transformation *over time* relates with changes in diabetes outcomes, we introduce as a predictor c_j^{diff}, the change in clinic score from 2008 to 2011: $c_j^{\text{diff}} \equiv c_j^{(2011)} - c_j^{(2008)}$. Clinics were de-identified and their corresponding scores were on an uninterpretable scale. However, the clinic score had a positive correlation with almost all questions in the PPCRS survey, and positive responses tend to represent a more mature transformation in primary care redesign. A simple histogram of $\boldsymbol{c}^{\text{diff}}$ reveals that all but one clinic score change is positive, implying that higher clinic score corresponds to a more mature primary

care transformation. Moreover, an analysis of the PPCRS data using Joint and Individual Variation Explained (JIVE)[24] only identified one latent component that was present across all CCM domains, and this component was closely correlated to the first principal component; this suggests that other structure in the PPCRS data is more granular. Further clustering analyses based on the Hamming distance[25] yielded only two groups of clinics, which can be thought of as having "more mature clinic transformation" and "less mature clinic transformation".

3.3.2 An Unadjusted Analysis

A simple correlation analysis between change in mean A1c score and $\boldsymbol{c}^{\mathbf{diff}}$ across clinics yields a significantly negative correlation: $\hat{\rho} = -0.318$ (p-value $= 0.001$). Figure 3.1 illustrates the marginal bivariate relationship between them with its simple linear regression line. These simple analyses suggest that a clinic that has made greater strides in primary care redesign observes lower mean A1c, a desirable diabetes outcome. However, this result considers only the marginal relationship between the the two variables, and fails to account for several important aspects, including (1) the hierarchical structure of the data, (2) potential dynamic confounders in the form of shifting clinic demographics, (3) the variability and heteroscedasticity of the outcome and covariates resulting from different clinic-year sample sizes, and (4) confounding between $\boldsymbol{c}^{\mathbf{diff}}$ and baseline clinic score $\boldsymbol{c}^{(\mathbf{2008})}$. As shown in later sections, a causal link between the change in clinic score and change in our outcome measures (including A1c score) is much less clear after accounting for these features.

3.3.3 Incorporating Health Care System Effects

Clinics within the same health care system are more similar in terms of resources, policies, and client demographics than clinics in different systems, so it is unreasonable to assume that intra-system clinics are uncorrelated. To explicitly model system-level dependence, we add system s-level effects υ_s and ψ_s to the likelihoods for $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\mathbf{diff}}$ in (3.1), respectively, in doing so adding an extra index s = 1, ..., S at the highest level of the hierarchy. The model in (3.1) can include these effects by adding system-specific parameters $\boldsymbol{\upsilon}$ and $\boldsymbol{\psi}$ to the mean structures for μ_{sj} and μ_{sj}^{diff}:

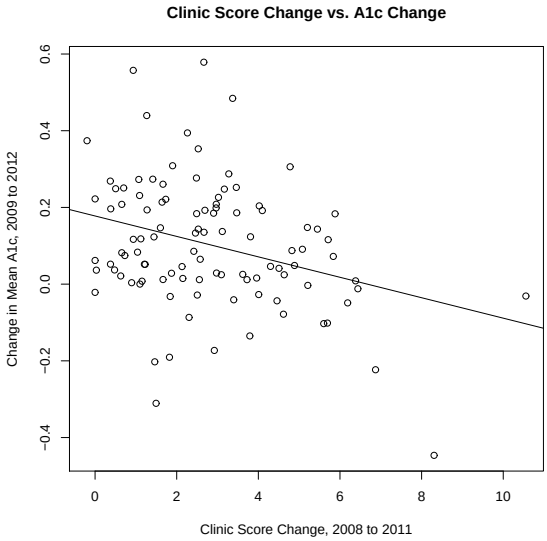


Figure 3.1: Changes in clinic score by changes in mean A1c with fitted simple linear regression line

$\mu_{sj} \sim N(T_{sj}\widetilde{\Delta}+\boldsymbol{X}_{sj}^T\widetilde{\boldsymbol{\beta}}+\upsilon_s, \widetilde{\tau}^2)$ and $\mu_{sj}^{\text{diff}} \sim N(T_{sj}\Delta+\boldsymbol{X}_{sj}^T\boldsymbol{\beta}+\psi_s, \tau^2)$. All system-specific effects are then assumed to arise from Gaussian distributions centered around 0: $\upsilon_s \overset{\text{iid}}{\sim} N(0,\omega^2)$; $\psi_s \overset{\text{iid}}{\sim} N(0,\zeta^2),\ s = 1, ..., S$. In the multivariate representation of (3.1), this is equivalent to modifying $\widetilde{\boldsymbol{\Omega}}$ and $\boldsymbol{\Omega}$ to be block diagonal matrices whose blocks correspond to systems. In $\widetilde{\boldsymbol{\Omega}}$, the diagonal elements are $\widetilde{\tau}^2+\omega^2$ and off-diagonal elements within blocks are ω^2. Similarly, in $\boldsymbol{\Omega}$, the diagonal elements are $\tau^2+\zeta^2$ and off-diagonal elements within blocks are ζ^2.

Recall that DiD estimation relies on the Parallel Trends assumption: clinics with a wealthier patient population may have better trends in diabetes outcomes prior to clinic restructuring, for example. Furthermore, each observation in model (3.1) is a clinic-year, which is the mean of patient-level data, so the estimate of Δ is subject to variation over time in the clinic's patient pool.[1] There could also be neighborhood level changes responsible for a clinic's demographic shifts. For example, gentrification may drastically change a local clinic's patient population, biasing our estimate of Δ if left unaccounted for. A patient's neighborhood's ACS-measured Wealth (Wlth) and Income/Education (Inc), as well as its percentage of non-Hispanic white residents (NHW), can be mapped to their MNCM record. Since not all patients at a given clinic live in the clinic's own ZIP code, we take

the clinic's sample mean of all patients' ZIP code Wlth, Inc, and NHW values as an estimate of a clinic's average neighborhood characteristics. These three patient-level residential characteristics are averaged within clinic and are included as candidate predictors in our analysis.

A clinic's trajectory in primary care transformation and mean diabetes outcomes may both be confounded by its access to resources, willingness and ability to restructure its primary care, and other factors present at the baseline year of 2008. Thus, the baseline clinic scores $\boldsymbol{c}^{(2008)}$ are included as a candidate predictor in the models for both $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\text{diff}}$ in (3.1). Accounting for this potential source of confounding also avoids "ceiling effects", where clinics that had already implemented many of the programs and resources measured by the PPCRS had consequently already improved diabetes outcomes, so that additional change spurred by health care home certification would be minimal. The interaction term $\boldsymbol{c}^{(2008)} * \boldsymbol{c}^{\text{diff}}$ was investigated as a potential covariate, but ultimately excluded as it did not have a clear effect nor did it improve model fit.

To address these potential sources of confounding, we include clinic-year-level changes in patient demographics meant to capture external phenomena driving changes in mean diabetes outcomes. We include as covariates the change in proportion of patients with commercial insurance, the change in mean age, the change in proportion of patients that were female, the change in proportion of patients with IVD, the change in average wealth, the change in average income/education, the change in average proportion of non-Hispanic white residents in the patients' ZIP codes, and baseline clinic score:

$$\boldsymbol{X}_{sj} = [1,\ c_{sj}^{\text{diff}},\ \text{Comm}_{sj}^{\text{diff}},\ \text{Age}_{sj}^{\text{diff}},\ \text{Fem}_{sj}^{\text{diff}},\ \text{IVD}_{sj}^{\text{diff}},\ \text{c}_{sj}^{(2008)},\ \text{Wlth}_{sj}^{\text{diff}},\ \text{Inc}_{sj}^{\text{diff}},\ \text{NHW}_{sj}^{\text{diff}}]$$

and $\boldsymbol{\beta} = [\beta_0,\ \Delta,\ \beta_1,\ ...,\ \beta_8]^T$.

3.3.4 Identifying Assumptions

In this subsection, we discuss Assumptions 1-3 in light of the MNCM data. Here, the "treatment" can be interpreted as a clinic increasing its speed of adopting the PCMH model by one clinic score unit higher from 2008 to 2011 (that is, $c_{sj}^{\text{diff}} = a + 1$ vs. $c_{sj}^{\text{diff}} = a$).

Model (3.1) conditions on "post-treatment" covariates, in that we are adjusting for the change

in covariates *after* a clinic adopts PCMH policies. There is an extensive literature warning against conditioning on post-treatment covariates when attempting to estimate a causal treatment effect,[26,27] in the clinical or observational setting; in our context, adjusting for post-treatment covariates may violate the assumption of Exogeneity (Assumption 1). Fortunately, the validity of this assumption is not a concern in our context. There is little empirical or contextual evidence to suggest that a clinic's speed to adopt the PCMH model affects clinic-level demographic changes; pairwise correlations between $\boldsymbol{c}^{\mathbf{diff}}$ and change in the seven covariates considered range from $\hat{\rho} = -0.20$ to $\hat{\rho} = 0.07$, with no associations significant at the p-value < 0.01 threshold. So, the adoption of the PCMH model has a small or non-existent effect on a patient's choice of clinic, a choice probably based more on infrequently changing variables like the patient's geographic location and insurance type. In the absence of Exogeneity, marginal structural models[28] represent an alternative approach that can still lead to consistent effect estimates.

In this context, Parallel Trends (Assumption 2) can be interpreted as if the actual observed c_{sj}^{diff} were a', and we substituted the counterfactual a, we would expect the same difference in outcome as if the observed value of c_{sj}^{diff} were a. While we are unable to empirically verify the appropriateness of Assumption 2, we have no contextual reasons to believe it is violated once we've adjusted for the covariates described in Section 3.3.3.

Because the primary care redesign legislation was passed in 2008, giving the clinics time between passage and the 2010 implementation to react, there is a risk that No Anticipatory Behavior (Assumption 3) is not met. The measure $\boldsymbol{c}^{\mathbf{diff}}$ is captured through the PPCRS, available only in 2008 and 2011, so we cannot examine Assumption 3 formally. However, in other work currently in progress, we use binary certification dates as our treatment variable. In that work, we compare trends in outcomes for this initial wave of certified clinics with later adopters. While we see some evidence of anticipatory behavior in later adopters, the "early wave" of clinics measured in this chapter shows no sign of anticipatory behavior. That is, pre-certification trends in outcomes among early wave clinics with larger $\boldsymbol{c}^{\mathbf{diff}}$ values were similar to those with smaller values.

3.3.5 Incorporating Error in Covariates

Model (3.1) assumes that its covariates, which are composed of clinic-year sample means and proportions and not patient-level data points, are fixed. This assumption does not capture the differing levels of variability in the mean covariates due to differing numbers of patient observations per clinic and differing levels of patient heterogeneity per clinic. To relax this assumption, we suppose that clinic j and year $^{(\text{yr})}$ Commercial insurance counts $\text{X.Comm}_{sj}^{(\text{yr})}$, mean age $\text{Age}_{sj}^{(\text{yr})}$, number of females $\text{X.Fem}_{sj}^{(\text{yr})}$, and number of patients with IVD $\text{X.IVD}_{sj}^{(\text{yr})}$ arise from binomial or Gaussian distributions with clinic j-specific parameters where appropriate; that is, we assume $\text{X.Comm}_{sj}^{(\text{yr})} \sim \text{Bin}\left(\text{n}_{sj}^{(\text{yr})}, \theta_{sj}^{(\text{yr})}\right)$, $\text{Age}_{sj}^{(\text{yr})} \sim N\left(\eta_{sj}^{(\text{yr})}, \sigma_{sj,\text{age}}^2/\text{n}_{sj}^{(\text{yr})}\right)$, $\text{X.Fem}_{sj}^{(\text{yr})} \sim \text{Bin}\left(\text{n}_{sj}^{(\text{yr})}, \gamma_{sj}^{(\text{yr})}\right)$, and $\text{X.IVD}_{sj}^{(\text{yr})} \sim \text{Bin}\left(\text{n}_{sj}^{(\text{yr})}, \kappa_{sj}^{(\text{yr})}\right)$, where $\text{Comm}_{sj}^{(\text{yr})} = \text{X.Comm}_{sj}^{(\text{yr})}/\text{n}_{sj}^{(\text{yr})}$, $\text{Fem}_{sj}^{(\text{yr})} = \text{X.Fem}_{sj}^{(\text{yr})}/\text{n}_{sj}^{(\text{yr})}$, and $\text{IVD}_{sj}^{(\text{yr})} = \text{X.IVD}_{sj}^{(\text{yr})}/\text{n}_{sj}^{(\text{yr})}$ with $\text{yr} \in \{2009, 2012\}$. We treat $c_{sj}^{(\text{yr})}$, $\text{Wlth}_{sj}^{(\text{yr})}$, $\text{Inc}_{sj}^{(\text{yr})}$, and $\text{NHW}_{sj}^{(\text{yr})}$ as fixed, as they are not patient-level data.

3.3.6 Prior Specification

Instead of imposing non-informative priors on each σ_{sj}^2 and $\sigma_{sj,\text{age}}^2$, we use a data-informed empirical Bayes approach to estimate inverse-gamma priors for each. The MNCM data are rich with patient-year observations at each clinic, so estimates of σ_{sj}^2 and $\sigma_{sj,\text{age}}^2$ are fairly accurate. We wish to use information across all clinics to infer the distribution of these variances. Since σ_{sj}^2 and $\sigma_{sj,\text{age}}^2$ are "level-one" parameters that describe the variance of patient data, we can simply match the shape and scale parameters with the first two moments of the inverse-gamma distribution. Namely, for $\pi(\sigma^2) \overset{\text{iid}}{\sim} \text{IG}(m, n),\ s = 1, ..., S,\ j = 1, ..., J_s$, the prior mean and variance are $E[\sigma^2] = n/(m-1)$ for $m > 1$ and $\text{Var}[\sigma^2] = n^2/[(m-1)^2(m-2)]$ for $m > 2$. Solving for m and n, we obtain

$$m = [E(\sigma^2)]^2/\text{Var}(\sigma^2) + 2 \ \text{ and } \ n = E(\sigma^2)\left\{[E(\sigma^2)]^2/\text{Var}(\sigma^2) + 1\right\}. \tag{3.3}$$

In our analysis, m and n were estimated by $\hat{m}$ and $\hat{n}$ by taking the clinic-level sample variances of the outcome $\boldsymbol{Y}$, $\hat{\boldsymbol{\sigma}^2} = [\hat{\sigma}_1^2, ...\hat{\sigma}_J^2]$, and substituting $E(\sigma_{sj}^2)$ and $\text{Var}(\sigma_{sj}^2)$ in (3.3) with $\hat{E}(\sigma_{sj}^2) =$

$(1/J)\sum_{s=1}^{S}\sum_{j\in s}\hat{\sigma}^2_{sj}$ and $\widehat{\text{Var}}(\sigma^2) = (1/(J-1))\sum_{s=1}^{S}\sum_{j\in s}(\hat{\sigma}^2_{sj} - \hat{E}(\sigma^2_{sj}))^2$, respectively. The prior for each σ^2_{sj} is then specified to be inverse-gamma with shape $\hat{m}$ and scale $\hat{n}$: $\pi(\sigma^2_{sj}) = \text{IG}(\hat{m}, \hat{n}),\ s = 1, ..., S,\ j = 1, ..., J_s$. The same process is used to estimate the prior distribution for each $\sigma^2_{sj,\text{age}}$, using the clinic-level sample variances of patient age: $\pi(\sigma^2_{sj,\text{age}}) = \text{IG}(\hat{m}_{\text{age}}, \hat{n}_{\text{age}})$, $s = 1, ..., S,\ j = 1, ..., J_s$.

We know little about the remaining distributions of the parameters in our model, so we use priors that are relatively uninformative. Specifically, we use flat priors for the regression coefficients, system-level hierarchical effects and clinic-specific mean age: $\pi(\widetilde{\boldsymbol{\beta}}) \propto \pi(\boldsymbol{\beta}) \propto \mathbf{1}$ and $\pi(\psi_s) \propto \pi(\upsilon_s) \propto \pi(\eta^{(\text{yr})}_{sj}) \propto 1$, uniform priors for proportion parameters: $\pi(\theta^{(\text{yr})}_{sj}) = \pi(\gamma^{(\text{yr})}_{sj}) = \pi(\kappa^{(\text{yr})}_{sj}) = U(0,1)$, and log-uniform priors for hierarchical variance parameters: $\pi(\tau^2) \propto 1/\tau^2$, $\pi(\widetilde{\tau}^2) \propto 1/\widetilde{\tau}^2$, $\pi(\zeta^2) \propto 1/\zeta^2$, $\pi(\omega^2) \propto 1/\omega^2$, each for $s = 1, ..., S$, $j = 1, ..., J_s$ and yr $\in \{2009, 2012\}$.

To test the sensitivity of our findings to the prior specification, we re-ran the analysis with the following priors: (i) quadrupled the variance while keeping the mean constant within each σ^2_{sj} and $\sigma^2_{sj,\text{age}}$ inverse-gamma prior, (ii) a Jeffreys prior[29] for each proportion parameter: $\pi(\theta^{(\text{yr})}_{sj}) = \pi(\gamma^{(\text{yr})}_{sj}) = \pi(\kappa^{(\text{yr})}_{sj}) = \text{Beta}(1/2, 1/2)$, and (iii) the remaining variance parameters: $\pi(\tau^2) \propto 1/\sqrt{\tau^2}$, $\pi(\lambda^2) \propto 1/\sqrt{\lambda^2}$, $\pi(\zeta^2) \propto 1/\sqrt{\zeta^2}$, $\pi(\omega^2) \propto 1/\sqrt{\omega^2}$. We found that the final results yielded by using these priors differ very little from those presented in Section 3.3.7.

3.3.7 Results

To avoid unnecessary complexity and collinearity (especially between highly correlated ACS covariates), a Deviance Information Criterion (DIC)[30] based forward selection algorithm was used to select covariates for the models for $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$ in (3.1). Specifically, the DIC is first computed with $\boldsymbol{X} = [\mathbf{1},\ \mathbf{c}^{\textbf{diff}}]$ (since Δ is our primary measure of interest, $\mathbf{c}^{\textbf{diff}}$ must be included in the model). Then, the reductions in DIC are computed for each predictor by adding only itself to $\boldsymbol{X}$, sorted by greatest reduction to least, added to $\boldsymbol{X}$ in order, and the algorithm stopped once the DIC is not reduced by the addition of another predictor.

For the change in A1c outcome, the forward variable selection yields the lowest DIC when $\boldsymbol{X} =$

$[\mathbf{1},\ \mathbf{c}^{\mathbf{diff}}, \mathbf{Comm}^{\mathbf{diff}}, \mathbf{Age}^{\mathbf{diff}}, \mathbf{Female}^{\mathbf{diff}}, \mathbf{IVD}^{\mathbf{diff}}, \mathbf{c}^{(\mathbf{2008})}]$, so that $\boldsymbol{\Theta} = [\beta_0, \Delta, \beta_1, \beta_2, \beta_3, \beta_4, \beta_5]^T$ is the full set of regression coefficients to estimate. For the change in LDL and change in SBP outcomes, the algorithm yields smaller subsets of this matrix, so we use the same covariate set across outcomes for simplicity. Posterior distributions were computed with 100,000 conditionally conjugate Gibbs draws[31] excluding the first 10,000 for burn-in, and convergence was confirmed with trace plots. The Gibbs sampling algorithm is presented in Appendix B.1. Table 3.1 displays, for each change in diabetes outcome, 95% equal-tail Bayesian credible intervals for each coefficient in $\boldsymbol{\Theta}$.

	Change in A1c	Change in LDL	Change in SBP
Change in clinic score, Δ	(-0.023, 0.013)	(-0.312, 0.558)	(-0.244, 0.186)
Change in prop. on commercial, β_1	(-0.157, 0.239)	(-3.885, 7.518)	(-1.711, 4.442)
Change in mean age, β_2	**(-0.314, -0.043)**	(-6.385, 1.990)	(-0.997, 3.058)
Change in prop. female, β_3	(-0.199, 0.310)	(-4.708, 9.848)	(-3.711, 4.224)
Change in prop. with IVD, β_4	(-0.315, 0.399)	(-10.169, 8.293)	(-6.290, 3.864)
Baseline clinic score, β_5	**(0.006, 0.031)**	(-0.186, 0.435)	(-0.251, 0.059)

Table 3.1: 95% Credible Intervals for Coefficients $\boldsymbol{\Theta}$

The posterior credible intervals (CI) for Δ include 0 across diabetes outcomes, suggesting that change in clinic score is not significantly associated with improving diabetes outcomes after adjusting for other changes in a clinic's patient characteristics and baseline clinic score. The change in mean age is negatively associated with A1c, suggesting that clinics whose patient pool got older on average saw an improvement in A1c score, adjusted for all other covariates. The baseline clinic score for A1c is positively significant, suggesting that clinics with greater baseline clinic score – indicating higher initial implementation of the programs and resources measured by the PPCRS – see patients whose A1c score worsened from 2008 to 2011. A robustness test removing the baseline clinic score results in an A1c 95% CI for Δ that is significantly negative, in contrast to the results above that include 0 in the A1c 95% CI for Δ.

The two significant credible intervals in Table 3.1 tell a surprising story: that clinics whose patient population got *older* and that had *fewer* PCMH-related resources at baseline observed *improved* mean A1c scores from 2009 to 2012. Since our analysis is limited to patients with type 2 diabetes,

this suggests that older patients living with diabetes control their A1c levels better than younger patients. There could also be survivorship bias present, in which the low-A1c patients living with diabetes survive to old age, while high-A1c younger patients do not. From a policy perspective, the significant positive signal for β_5 suggests a ceiling effect of PCMH clinics – that is, those clinics which start out with many PCMH-related resources had smaller opportunity for improvement in diabetes outcomes than clinics with few PCMH-related resources. A brief analysis revealed that clinics in the top $\mathbf{c}^{(2008)}$ quartile started with low mean A1c scores, and these scores increased from 2009 to 2012, while clinics in the bottom $\mathbf{c}^{(2008)}$ quartile had unchanging high mean A1c scores. This suggests that clinics with more baseline access to resources serve communities more representative of the broader population that is worsening over time with respect to diabetes outcomes,[32] while clinics with less baseline access to resources serve communities that fail to control A1c levels consistently across time.

In the unadjusted analysis of Section 3.3.2, clinic-year sample sizes were ignored, allowing for patient outcomes at smaller clinics to influence the results more. Figure 3.2 illustrates how accounting for a clinic's size and changes in covariates shrinks the sample difference in mean A1c ($\bar{\boldsymbol{Y}}^{(2012)} - \bar{\boldsymbol{Y}}^{(2009)}$) towards their predicted values $\boldsymbol{X}\hat{\boldsymbol{\Theta}}$ via their latent clinic effects $\boldsymbol{\mu}^{\text{diff}}$, where $\hat{\boldsymbol{\Theta}}$ is the posterior mean of $\boldsymbol{\Theta}$. In Figure 3.2, the sample difference of clinic sj was scaled by $1/(1/n_{sj}^{(2009)} + 1/n_{sj}^{(2012)})$, so that larger clinics appear relatively larger than smaller clinics. Closed circles correspond to the clinic's predicted mean change $\boldsymbol{\mu}^{\text{diff}}$. Clinics belonging to the three largest systems are colored magenta, blue, and green. On average, smaller clinics are shrunk significantly towards their predicted values, while estimates of sample difference in mean A1c at larger clinics are relatively precise and shrunk much less.

Similar plots showing shrinkage for the difference in mean LDL and mean SBP are shown in Figure 3.3. For these outcomes the predictive model is less strong and dominated by the error terms, and thus the plots show no clear trend. However, both plots show substantial clustering at the system-level, and the primary mode of shrinkage appears to be toward the system effects rather than toward the predicted values.

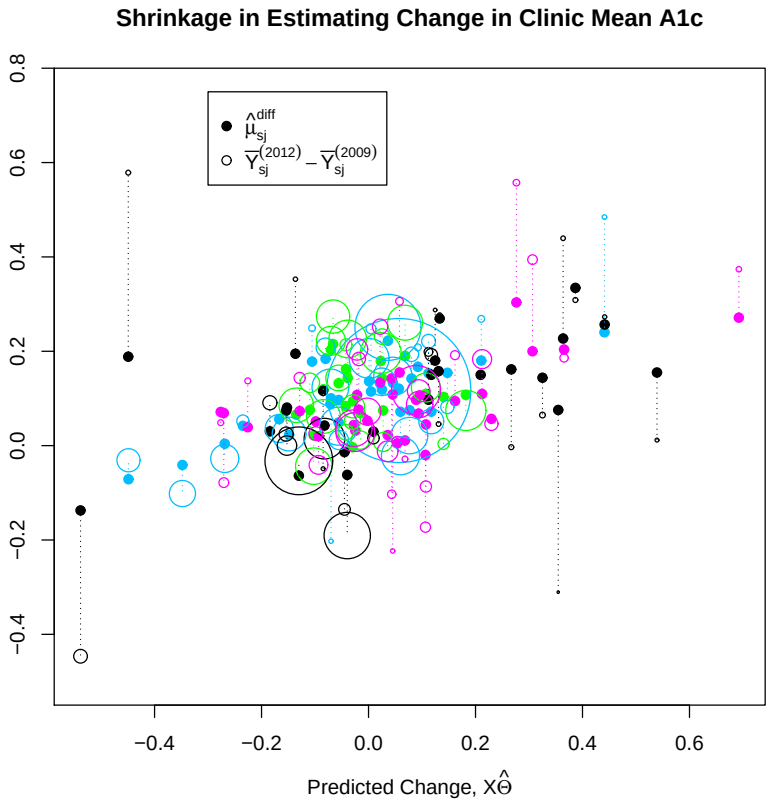


Figure 3.2: Shrinkage of clinics' sample differences towards their predicted differences given by $\boldsymbol{X}\hat{\boldsymbol{\Theta}}$ for A1c, scaled by their relative sample sizes; clinics within the three largest systems are colored magenta, blue, and green.

3.3.8 Discussion, Limitations, and Future Directions

In this section we identified one joint component across all question domains for the 2008 and 2011 PPCRS data, and considered whether a change in this component was causally associated with a change in diabetes outcomes. Although we did not find strong evidence of an effect for this component, it is perhaps an over-simplified measure of PCMH-related programs and resources at the clinic level. Additionally, estimation of the causal effect using our HDiD model (and DiD models in general) depends on the timing of the post-treatment analysis; 2012 may be too soon for any true causal relationship to have fully manifested. In Chapter 4, we apply a similar analysis (with a few modifications) using 2017 as the post-treatment timepoint, and find that $\boldsymbol{c}^{\mathbf{diff}} = \boldsymbol{c}^{(\mathbf{2017})} - \boldsymbol{c}^{(\mathbf{2008})}$ is significantly related to improved LDL and SBP. We also stress that there still may be

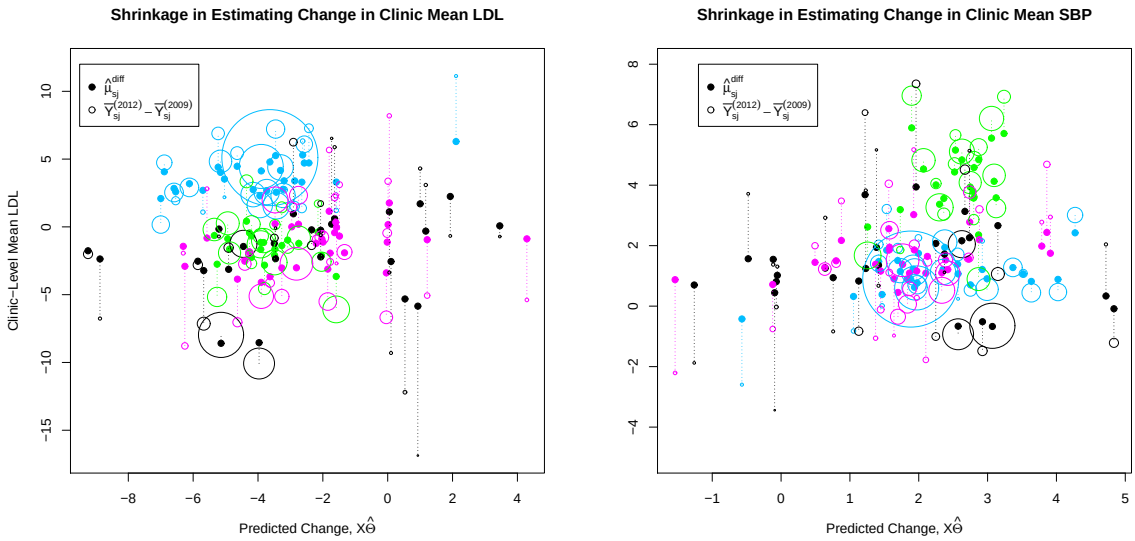


Figure 3.3: Shrinkage of clinics' sample differences for LDL (left) and SBP (right), scaled by their relative sample sizes; clinics within the three largest systems are colored magenta, blue, and green.

specific question domains, and policy changes more generally, that are causally relevant to improved diabetes outcomes. We believe that analyses of later PPCRS data will uncover more specific factors as important components that are relevant to diabetes care. In fact, a recent principal components analysis of the PPCRS sent out to a broader sample of clinics in 2017 shows a much richer structure, where up to ten groups of questions arise as important sources of survey answer variation. This motivates future work to uncover which factors are most important in diabetes care in the primary care setting, as panel data emerge for this larger sample.

Improvement in A1c should not really be measured linearly across its entire scale; for example, a patient A1c level less than 6.5% may be considered satisfactory and not in need of further reduction. However, we are aggregating patient-year level A1c scores to a clinic-year mean (each with at least 30 patient-years), and the difference in means across should still provide a reasonable summary of clinic-level A1c improvement. Encryption of patient identifiers was not consistent across years, so any inference about the effectiveness of the clinic transformation was not possible at the patient level, so our clinic-level analysis clearly lacks precision. Also, there is a potential for recall bias in the PPCRS survey data, since the survey was implemented in 2011 and responses for 2008 were

based on recall. All c_{sj}^{diff} were non-negative except for one, evidence that improved programs and resources measured from 2008 to 2011 may be artificially created by PPCRS respondents' recall. In fact, a preliminary analysis of PPCRS data from 2011 and 2017 shows that some clinics in fact regress in maturity of primary care transformation, further evidence that recall bias may be present in the 2008 to 2011 analysis. Future work will investigate which *particular* items of PPCRS are related to the biggest changes in diabetes outcomes.

Our model assumes a homogenous treatment effect; that is, each patient visiting a clinic whose c_{sj}^{diff} value had been one unit higher benefits by a single value (Δ). We could relax this assumption by developing our model within a more general "Changes-in-Changes" (CiC) model framework, which assumes the potential outcomes are functions $h(u, t)$ increasing monotonically in a patient's unobservable characteristics u. The CiC model is much more general than the DiD model, as it allows for differential treatment effects across individuals and groups, has identifying assumptions which are invariant to monotone transformations, and allows for discrete outcomes and more than two timepoints, all while enjoying asymptotic normality and consistency properties under mild conditions.[33]

The discrepancy between the results of the unadjusted analysis in Section 3.3.2 and Table 3.1 should serve as a cautionary tale for investigators. Failing to account for confounding and all sources of variability – system structure, covariates and their variability, baseline clinic score – could have lead an investigator to falsely conclude a clear association between improvements in A1c score and $\boldsymbol{c}^{\text{diff}}$.

Chapter 4

Bayesian Variable Selection in Hierarchical Difference-in-Differences Models

The HDiD model introduced in this work adjusts the baseline and change models for group-level covariates, but has thus far offered no justification why such covariate adjustment is necessary. Since Δ measures the covariate-adjusted relationship between the treatment and $\boldsymbol{\mu}^{\mathbf{diff}}$, it may not be intuitive why the model in (3.1) includes a covariate-adjusted baseline model. Even if the change model is appropriately specified, we show that estimates of Δ may still be biased if predictors correlate with $\boldsymbol{\mu}$, $\boldsymbol{\mu}^{\mathbf{diff}}$, and the treatment but are not included in the baseline model.[17] Further, it's not immediately clear how to decide which covariates to include in the HDiD context. In this chapter, we review some literature relevant to variable selection in causal inference and difference-in-differences, establish theoretical and empirical results characterizing the role covariate adjustment plays in estimating Δ, and then propose four variable selection methods in the Bayesian HDiD model framework.

4.1 Variable Selection in Causal Inference

Adjusting an estimator of a causal treatment effect to eliminate confounding is a well-explored topic in the single timepoint context. Propensity score matching and inverse probability of treatment weighting are popular methods which use conditional exchangeability (i.e., conditional on some other covariate(s), treatment groups are comparable) to eliminate confounding.[34,35] However, these methods require adjustment for *all* confounding variables to estimate the treatment effect consistently.

More recently, approaches which jointly estimate outcome and exposure/treatment models, and then leverage information across each model to estimate the treatment effect have been proposed. Wang et al.[36] and Cefalu[37] propose methods which identify potential confounders by imposing prior dependence on a covariate's inclusion in the propensity score and outcome models. Zigler and Dominici propose Bayesian methods that perform variable selection, then estimate the treatment effect as a weighted average of estimates yielded from propensity score models with different covariates included.[38] Koch et al. propose a method which simultaneously estimates the treatment effect while performing an adaptive group lasso based variable selection algorithm.[39]

In the point-exposure context, confounding occurs when a covariate which is associated with both treatment and outcome is not accounted for in the treatment-outcome analysis. In the DiD context, confounding occurs when (i) the covariate is associated with treatment and (iia) the association between the covariate and outcome varies over time or (iib) the covariate evolves over time differently in the treatment and control groups.[40] Thus, identifying the causal effect Δ may require extending model (1.1) as follows:

$$Y_i^{(t)} = \beta_0(1 - T_i) + \beta_0^{'} T_i + \phi \mathbb{1}(t = 1) + \Delta T_i \mathbb{1}(t = 1) + \sum_{k=1}^{K} \beta_k^{(t)} X_{ik}^{(t)} + \epsilon_i^{(t)}. \tag{4.1}$$

Static covariates ($X_{ik}^{(0)} = X_{ik}^{(1)}$) may still be confounders of type (iia) if $\beta_k^{(0)} \neq \beta_k^{(1)}$; confounders

of type (iib) require $X_{ik}^{(0)} \neq X_{ik}^{(1)}$. We can similarly extend model (1.2) to include covariates:

$$Y_i^{\text{diff}} = \phi + \Delta T_i + \sum_{k=1}^{K} \beta_k X_{ik} + \epsilon_i \tag{4.2}$$

For static covariates (i.e., $X_{ik} = X_{ik}^{(1)} = X_{ik}^{(0)}$ so that $\beta_k = \beta_k^{(1)} - \beta_k^{(0)}$) and dynamic covariates with constant association with the outcome (i.e., $X_{ik} = X_{ik}^{(1)} - X_{ik}^{(0)}$ so that $\beta_k = \beta_k^{(1)} = \beta_k^{(0)}$), this is a classical linear model, in which numerous frequentist variable selection techniques exist. There also exist numerous Bayesian techniques for performing model selection, such as computing Bayes Factors (BF) or DIC (the latter of which we used in the forward selection algorithm in Section 3.3.7). The major limitation of these two Bayesian methods is, for large number of predictors K, computing 2^K models' BFs or DICs is wildly inefficient. However, with a clever prior choice for coefficients $\boldsymbol{\beta}$, one can leverage a conditionally conjugate Gibbs sampling algorithm which automatically identifies which predictors to include and exclude.

Many applications of the covariate-adjusted DiD model directly select these covariates without any data-driven variable selection techniques.[41–43] Stuart et al. uses propensity scores to increase comparability of the four DiD groups (treatment-pre, control-pre, treatment-post, control-post).[44] Sofer et al. interpret the DiD model as a negative control outcome to identify and correct for biased DiD-estimated treatment effects.[45]

4.2 Covariate Adjustment in Hierarchical Difference-in-Differences Models

4.2.1 Bias Due to Misspecified Baseline Model

The main objective of the HDiD model is to estimate the treatment effect, so it may be tempting to focus solely on covariate adjustment in the change model. In (3.1), we adjust the estimates of $\boldsymbol{\mu}$ with the same covariate set as $\boldsymbol{\mu}^{\text{diff}}$, a modeling choice that may seem like an unnecessary over-complication. In this section, we investigate the bias incurred upon misspecification of the

baseline model, and go on in the next sections to present general results capturing the bias incurred in estimating $\boldsymbol{\Theta} = [\widetilde{\Delta},\ \widetilde{\boldsymbol{\beta}}^T,\ \Delta,\ \boldsymbol{\beta}^T]^T$ when important covariates are not included.

Even if the predictive model for $\boldsymbol{\mu}^{\textbf{diff}}$ is appropriately specified, estimates of Δ may still be biased if predictors correlate with $\boldsymbol{\mu}$ but are not included in the model for $\boldsymbol{\mu}$ in (3.1). We conduct a simple simulation study to illustrate this by comparing the bias incurred when estimating Δ and $\boldsymbol{\beta}$ under two different mean structures for $\boldsymbol{\mu}$: $E[\boldsymbol{\mu}|\mu_0] = \mu_0\mathbf{1}$ (MS1) and $E[\boldsymbol{\mu}|\boldsymbol{X}] = \boldsymbol{X}\widetilde{\boldsymbol{\Theta}}$ (MS2), where $\mathbf{1}$ is a vector of 1's.

First, treatment status $\boldsymbol{T}$ and covariates $\boldsymbol{X_1}, \boldsymbol{X_2}$, and $\boldsymbol{X_3}$ are generated from independent and identically distributed multivariate normal distributions: $\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3} \overset{iid}{\sim} \boldsymbol{N}_J(\mathbf{0}, \boldsymbol{I}_J)$, so that $\boldsymbol{X} = [\boldsymbol{X_1}, \boldsymbol{X_2}, \boldsymbol{X_3}]$. We then generate $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$ according to multivariate normal distributions with uncorrelated errors: $\boldsymbol{\mu} \sim \boldsymbol{N}_J(\boldsymbol{T}\widetilde{\Delta} + \boldsymbol{X}\widetilde{\boldsymbol{\beta}}, \boldsymbol{I}_J)$ and $\boldsymbol{\mu}^{\textbf{diff}} \sim \boldsymbol{N}_J(\boldsymbol{T}\Delta + \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{I}_J)$, with $\widetilde{\Delta} = 1$, $\Delta = 0$, $\widetilde{\boldsymbol{\beta}} = [0.5, 0, 0]^T$ and $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]^T = [0.5, 1, 0]^T$. Thus, $\boldsymbol{T}$ is predictive of only $\boldsymbol{\mu}$, $\boldsymbol{X}_2$ is predictive of only $\boldsymbol{\mu}^{\textbf{diff}}$, and $\boldsymbol{X}_1$ is predictive of both. Individual level outcome data for pre- and post-treatment are then generated from Gaussian distributions according to the model in (3.1): $Y_{ji}^{(0)} \sim N(\mu_j, 1)$ and $Y_{ji}^{(1)} \sim N(\mu_j + \mu_j^{\text{diff}}, 1)$, for subjects $i = 1, ..., n_j$, where $n_j = 10$ for each group $j = 1, ..., J$, where $J = 100$. In both scenarios, $\boldsymbol{\mu}^{\textbf{diff}}$ was modeled correctly as a Gaussian with mean $\boldsymbol{T}\Delta + \boldsymbol{X}\boldsymbol{\beta}$: $\boldsymbol{\mu}^{\textbf{diff}} \sim \boldsymbol{N}_J(\boldsymbol{T}\Delta + \boldsymbol{X}\boldsymbol{\beta}, \boldsymbol{I}_J)$. We modeled $\boldsymbol{\mu}$ under two different approaches: as a random effect with no covariates, $\boldsymbol{\mu} \sim \boldsymbol{N}_J(\mu_0\mathbf{1}_J, \boldsymbol{I}_J)$ (MS1), and another with the same predictors used in the model for $\boldsymbol{\mu}^{\textbf{diff}}$, $\boldsymbol{\mu} \sim \boldsymbol{N}_J(\boldsymbol{T}\widetilde{\Delta} + \boldsymbol{X}\widetilde{\boldsymbol{\beta}}, \boldsymbol{I}_J)$ (MS2), where $\mathbf{1}$ is a vector of 1's and $\boldsymbol{I}$ is the identity matrix. A simulation was conducted with 10,000 replications, each using 10,000 conditionally conjugate Gibbs draws excluding the first 1,000 for burn-in. The full conditionals are simpler cases of those presented in Appendix B, so the Gibbs sampling algorithm is omitted from this work. Each parameter was initialized to its true value, and convergence was confirmed via trace plots.

Table 4.1 displays the estimated biases incurred in estimating Δ and $\boldsymbol{\beta}$, to predict $\boldsymbol{\mu}^{\textbf{diff}}$, under models MS1 or MS2 for $\boldsymbol{\mu}$. Precisely those nonzero coefficients that MS1 does not account for lead to substantial bias in the corresponding coefficients: $\widehat{\text{Bias}}(\hat{\Delta})$ = 0.089 and $\widehat{\text{Bias}}(\hat{\beta}_1)$ = 0.046.

	$\widehat{\text{Bias}}(\hat{\Delta})$	$\widehat{\text{Bias}}(\hat{\beta}_1)$	$\widehat{\text{Bias}}(\hat{\beta}_2)$	$\widehat{\text{Bias}}(\hat{\beta}_3)$
MS1	0.089	0.046	-0.002	0.002
MS2	-0.002	0.000	-0.002	0.002

Table 4.1: Estimated Biases ($\widehat{\text{Bias}}$) of Δ and $\boldsymbol{\beta}$ Under Different Mean Structures (MS) for $\boldsymbol{\mu}$

Those regression coefficients in $\boldsymbol{\beta}$ corresponding to truly zero coefficients in $\widetilde{\boldsymbol{\beta}}$ incur virtually no bias, which is sensible since the columns of $\boldsymbol{X}$ are independent and MS1 does not include $\boldsymbol{X_2}$ and $\boldsymbol{X_3}$. Using MS2 incurs virtually no bias in estimating Δ or $\boldsymbol{\beta}$, even though the model for $\boldsymbol{\mu}^{\textbf{diff}}$ was the same, illustrating that $\boldsymbol{\mu}$ must share these predictors to avoid biased estimates of Δ and β_1.

4.2.2 Causal Relationships of Interest

	$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$	$\boldsymbol{X}_4$	$\boldsymbol{X}_5$	$\boldsymbol{X}_6$	$\boldsymbol{X}_7$	$\boldsymbol{X}_8$
$\boldsymbol{T}$	✓	✓	✓	✓				
$\boldsymbol{\mu}$	✓	✓			✓	✓		
$\boldsymbol{\mu}^{\textbf{diff}}$	✓		✓		✓		✓	

Table 4.2: Covariates and their role in the causal graph

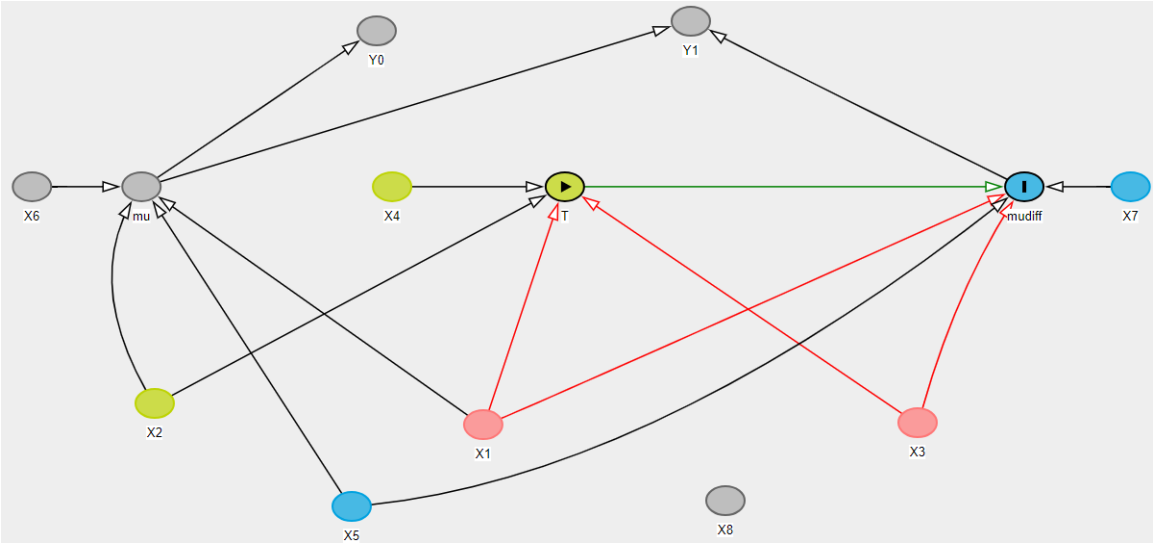


Figure 4.1: Assumed causal relationships and data generation throughout the chapter

In order to further motivate the need for covariate adjustment and variable selection techniques, we introduce a simple example of the assumed data generation in Equation (3.1). The observed subject-level pre-treatment outcome $Y_{ji}^{(0)}$ is assumed to arise from a group j-specific pre-treatment

mean μ_j and the observed subject-level post-treatment outcome $Y_{ji}^{(1)}$ is assumed to arise from a group j-specific pre-treatment mean μ_j modified by a group j-specific mean change μ_j^{diff}. It is then assumed that μ_j and μ_j^{diff} arise from a mean structure modified by covariates through the design matrix $\boldsymbol{X}$ and treatment vector $\boldsymbol{T}$; this includes the treatment $\boldsymbol{T}$ possibly affecting $\boldsymbol{\mu}^{\textbf{diff}}$, as well as variables which could affect any combination of $\boldsymbol{T}$, $\boldsymbol{\mu}^{\textbf{diff}}$, and $\boldsymbol{\mu}$.

For our theoretical and simulation evaluations, we consider a hypothetical set of covariates that cover each of the $2^3 = 8$ combinations of does/does not affect $\boldsymbol{T}$, does/does not affect $\boldsymbol{\mu}$, and does/does not affect $\boldsymbol{\mu}^{\textbf{diff}}$; Table 4.2 lists these cases. Additionally, Figure 4.1 displays the assumed data generation process as a directed acyclic graph,[46] and was generated using the "DAGitty" R package.[47] Our goal is to estimate the causal effect of $\boldsymbol{T}$ on $\boldsymbol{\mu}^{\textbf{diff}}$, represented by the green line in Figure 4.1.

4.2.3 Data Generation

Throughout this chapter, we use simulation to assess the operating characteristics of our approaches. Below, we describe the data generation used in each simulation.

1. Generate $\boldsymbol{X}_1, ..., \boldsymbol{X}_K \overset{iid}{\sim} \boldsymbol{N_J}(\boldsymbol{0}, \boldsymbol{I_J})$, where $\boldsymbol{I_J}$ is the J-dimensional identity matrix. Let $\boldsymbol{X} = [\boldsymbol{X}_1, ..., \boldsymbol{X}_K]$.

2. Generate $\boldsymbol{T} \sim \boldsymbol{N_J}(\boldsymbol{X\alpha}, \boldsymbol{I_J})$.

3. Generate $\boldsymbol{\mu} \sim \boldsymbol{N_J}(\boldsymbol{X}\widetilde{\boldsymbol{\beta}}, \boldsymbol{I_J})$ and $\boldsymbol{\mu}^{\textbf{diff}} \sim \boldsymbol{N_J}(\boldsymbol{T\Delta} + \boldsymbol{X\beta}, \boldsymbol{I_J})$.

4. Generate $Y_{ji}^{(0)} \sim N(\mu_j, 1)$ and $Y_{ji}^{(1)} \sim N(\mu_j + \mu_j^{\text{diff}}, 1)$, for $j = 1, ..., J$ and $i = 1, ..., n_j^{(t)}$.

So, $\boldsymbol{\alpha}$ controls which variables are predictive of treatment, $\widetilde{\boldsymbol{\beta}}$ controls which variables are predictive of baseline mean, and $\boldsymbol{\beta}$ controls which variables are predictive of change in means. For example, a generative model that includes all eight hypothetical covariates in Table 4.2 and Figure 4.1 is given by setting $\boldsymbol{\alpha} = [1, 1, 1, 1, 0, 0, 0, 0]^T$, $\widetilde{\boldsymbol{\beta}} = [1, 1, 0, 0, 1, 1, 0, 0]^T$, $\boldsymbol{\beta} = [1, 0, 1, 0, 1, 0, 1, 0]^T$,

$\widetilde{\Delta} = 0$, and $\Delta = 1$. For each simulation, we use either $J = 50$ or $J = 100$ groups, each having $n_j^{(0)} = n_j^{(1)} = 10$ subjects, $j = 1, ..., J$.

4.2.4 Omitted Variable Bias

After marginalizing $\boldsymbol{Y}^{(0)}$ and $\boldsymbol{Y}^{(1)}$ over $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$ and then concatenating them into a vector $\boldsymbol{Y}$, the model in (3.1) can be re-expressed as $\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{AB_1\Theta_1, \Sigma})$, where $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{A_0} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{A_1} \end{bmatrix}$ assigns pre- and post-treatment group-level predictors to the subject level, $\boldsymbol{B_1} = \begin{bmatrix} \boldsymbol{X}_{\widetilde{w}=1} & \widetilde{\boldsymbol{T}} & \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{X}_{\widetilde{w}=1} & \widetilde{\boldsymbol{T}} & \boldsymbol{X}_{w=1} & \boldsymbol{T} \end{bmatrix}$ where subscripts $_{\widetilde{w}=1}$ and $_{w=1}$ subset the object to those elements included in the baseline and change models (respectively), and $\widetilde{\boldsymbol{T}}$ is $\boldsymbol{T}$ when $\boldsymbol{T}$ is included as a predictor in the baseline model and $\boldsymbol{0}$ when it is not, and $\boldsymbol{\Theta_1} = [\widetilde{\boldsymbol{\beta}}_{\widetilde{w}=1}\ \widetilde{\Delta}\ \boldsymbol{\beta}_{w=1}\ \Delta]^T$. Whereas $\{\boldsymbol{B_1}, \boldsymbol{\Theta}_1\}$ define the model used for estimation, assume the true generative model is $\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{AB\Theta, \Sigma})$ where $\boldsymbol{B} = [\boldsymbol{B_1}\ \boldsymbol{B_0}]$ with $\boldsymbol{B_0} = \begin{bmatrix} \widetilde{\boldsymbol{U}} & \boldsymbol{0} \\ \widetilde{\boldsymbol{U}} & \boldsymbol{U} \end{bmatrix}$ such that $\widetilde{\boldsymbol{U}}$ and $\boldsymbol{U}$ are those covariates (with $\widetilde{\boldsymbol{U}}$ including $\boldsymbol{T}$ when not adjusting $\boldsymbol{\mu}$ for $\boldsymbol{T}$) excluded from the baseline and change models (respectively), and $\boldsymbol{\Theta} = [\boldsymbol{\Theta}_1\ \boldsymbol{\Theta}_0]^T$ where $\boldsymbol{\Theta}_0$ is a vector of coefficients corresponding to those covariates excluded from the model. We can then apply a modified version of the familiar omitted-variable bias result:[48]

Theorem 2. *Suppose $\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{AB\Theta}, \boldsymbol{\Sigma})$, where $\boldsymbol{B}$ is a matrix and $\boldsymbol{\Theta}$ is a vector, each which can be partitioned into $[\boldsymbol{B_1}\ \boldsymbol{B_0}]$ and $[\boldsymbol{\Theta}_1\ \boldsymbol{\Theta}_0]^T$ respectively, and $\boldsymbol{A}$ is a matrix of conforming dimension. Suppose $\boldsymbol{Y}$ is modeled as $\boldsymbol{Y} \sim \boldsymbol{N}(\boldsymbol{AB_1\Theta}_1, \boldsymbol{\Sigma})$. Finally, let $\boldsymbol{\Theta}_1$ have a flat prior. Then, the bias of $\widehat{\boldsymbol{\Theta}}_1$ as estimated by the model is*

$\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\Theta}}[\widehat{\boldsymbol{\Theta}}_1 - \boldsymbol{\Theta}_1] = (\boldsymbol{B_1}^T\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{AB_1})^{-1}\boldsymbol{B_1}^T\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{AB_0\Theta}_0$.

Proof. The posterior distribution of $\widehat{\boldsymbol{\Theta}}_1$ is

$$p(\widehat{\boldsymbol{\Theta}}_1|\boldsymbol{Y}) = \boldsymbol{N}(\boldsymbol{VB_1}^T\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{Y}, \boldsymbol{V}),\ \text{where}\ \boldsymbol{V} = (\boldsymbol{B_1}^T\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{AB_1})^{-1} \tag{4.3}$$

Then, $\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\Theta}}[\widehat{\boldsymbol{\Theta}}_1 - \boldsymbol{\Theta}_1]$

$= \boldsymbol{V}\boldsymbol{B_1}^T\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\mathbb{E}_{\boldsymbol{Y}|\boldsymbol{\Theta}}\boldsymbol{Y} - \boldsymbol{\Theta}_1$

$= \boldsymbol{V}\boldsymbol{B_1}^T\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}(\boldsymbol{B_1}\boldsymbol{\Theta}_1 + \boldsymbol{B_0}\boldsymbol{\Theta}_0) - \boldsymbol{\Theta}_1$

$= (\boldsymbol{B_1}^T\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{B_1})^{-1}\boldsymbol{B_1}^T\boldsymbol{A}^T\boldsymbol{\Sigma}^{-1}\boldsymbol{A}\boldsymbol{B_0}\boldsymbol{\Theta}_0$ ■

In our context, $\boldsymbol{\Theta}_1 = [\widetilde{\boldsymbol{\beta}}_{\widetilde{\boldsymbol{w}}=1}\ \widetilde{\Delta}_1\ \boldsymbol{\beta}_{\boldsymbol{w}=1}\ \Delta]^T$ and $\boldsymbol{\Theta}_0 = [\widetilde{\boldsymbol{\beta}}_{\widetilde{\boldsymbol{w}}=0}\ \widetilde{\Delta}_0\ \boldsymbol{\beta}_{\boldsymbol{w}=0}]^T$ where $\widetilde{\Delta}_1 = \widetilde{\Delta}$ when $\boldsymbol{T}$ is included as a predictor in the baseline model and 0 otherwise and $\widetilde{\Delta}_0 = \widetilde{\Delta}$ when $\boldsymbol{T}$ is excluded as a predictor from the baseline model and 0 otherwise, and $\boldsymbol{\Sigma} = \begin{bmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{bmatrix}$, where $\boldsymbol{\Sigma}_{11}$ is a block diagonal matrix where the j'th block is a diagonal matrix with $\widetilde{\sigma}_j + \widetilde{\tau}^2$ on the diagonal and $\widetilde{\tau}^2$ on the within-block off-diagonals, $\boldsymbol{\Sigma}_{12}$ and $\boldsymbol{\Sigma}_{21}$ are block diagonal matrices with every element in each block being $\widetilde{\tau}^2$, and $\boldsymbol{\Sigma}_{22}$ is a block diagonal matrix where the j'th block is a diagonal matrix with $\sigma_j^2 + \tau^2 + \widetilde{\tau}^2$ on the diagonal and $\tau^2 + \widetilde{\tau}^2$ on the within-block off-diagonals.

Isolating the expression in (4.3) to just Δ is very difficult analytically, so we perform a brief simulation to see which covariates we need to include when we sequentially add predictors to both $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$, comparing the bias of $\widehat{\Delta}$ when we (1) do not adjust the baseline model for $\boldsymbol{T}$, and (2) adjust the baseline model for $\boldsymbol{T}$. The simulations proceed by, for each iteration, generating data via Section 4.2.3 and computing the bias result above.

	No adjust for $\boldsymbol{T}$	Adjust for $\boldsymbol{T}$
Null	0.437	0.400
+ X_1	0.271	0.250
+ X_3	0.031	-0.001
+ X_2	0.002	0.000
+ X_5	0.000	0.002
+ X_7	0.000	0.000
+ X_6	0.000	0.000
+ X_4	0.000	0.000
Full	0.000	0.000

Table 4.3: Bias of $\widehat{\Delta}$ by including the current and all preceding rows as covariates for $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$, without and with adjusting $\boldsymbol{\mu}$ for $\boldsymbol{T}$

Table 4.3 shows the results of this simulation, where each row correponds to the addition of a new covariate into the baseline and change models and the columns correspond to whether or not we adjust $\boldsymbol{\mu}$ for $\boldsymbol{T}$. As an example, the "+ X_3" and "Adjust for $\boldsymbol{T}$" cell adjusts the model for $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$

with $\{\boldsymbol{T}, \boldsymbol{X}_1, \boldsymbol{X}_3\}$ as covariates. The results show that when we do not adjust $\boldsymbol{\mu}$ for $\boldsymbol{T}$, we must adjust $\boldsymbol{\mu}$ for all covariates predictive of both $\boldsymbol{\mu}$ and $\boldsymbol{T}$ *and* adjust $\boldsymbol{\mu}^{\textbf{diff}}$ for all covariates predictive of both $\boldsymbol{\mu}^{\textbf{diff}}$ and $\boldsymbol{T}$ (here, $\{\boldsymbol{X}_1, \boldsymbol{X}_2, \boldsymbol{X}_3\}$) to achieve unbiasedness in estimating Δ. When we do adjust $\boldsymbol{\mu}$ for $\boldsymbol{T}$, Table 4.3 suggests that we only need to adjust for all covariates predictive of $\boldsymbol{T}$ and $\boldsymbol{\mu}^{\textbf{diff}}$ (here, $\{\boldsymbol{X}_1, \boldsymbol{X}_3\}$) to acheive unbiasedness.

4.2.5 Variable-by-Variable Performance

To isolate the impact of each combination of does/does not affect $\boldsymbol{T}$, does/does not affect $\boldsymbol{\mu}$, and does/does not affect $\boldsymbol{\mu}^{\textbf{diff}}$ has on the estimation of Δ, for various modeling choices, we conduct a simulation. We define eight data generation scenarios by the process described in Section 4.2.3; each scenario uses only one of the covariates described in Table 4.2 as the single true generating covariate and (depending on the modeling choice) the single model covariate. Then, we estimate Δ according to each of the following eight model choices:

- Choice 1: Do not adjust $\boldsymbol{\mu}$ or $\boldsymbol{\mu}^{\textbf{diff}}$ for $\boldsymbol{X}_k$.
- Choice 2: Adjust $\boldsymbol{\mu}$, but do not adjust $\boldsymbol{\mu}^{\textbf{diff}}$, for $\boldsymbol{X}_k$.
- Choice 3: Adjust $\boldsymbol{\mu}^{\textbf{diff}}$, but do not adjust $\boldsymbol{\mu}$, for $\boldsymbol{X}_k$.
- Choice 4: Adjust $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$ for $\boldsymbol{X}_k$.
- Choice 5: Adjust $\boldsymbol{\mu}$ for $\boldsymbol{T}$, do not adjust $\boldsymbol{\mu}^{\textbf{diff}}$ for $\boldsymbol{X}_k$.
- Choice 6: Adjust $\boldsymbol{\mu}$ for $\boldsymbol{T}$ and $\boldsymbol{X}_k$, but do not adjust $\boldsymbol{\mu}^{\textbf{diff}}$, for $\boldsymbol{X}_k$.
- Choice 7: Adjust $\boldsymbol{\mu}$ for $\boldsymbol{T}$, adjust $\boldsymbol{\mu}^{\textbf{diff}}$ for $\boldsymbol{X}_k$.
- Choice 8: Adjust $\boldsymbol{\mu}$ for $\boldsymbol{T}$ and $\boldsymbol{X}_k$, adjust $\boldsymbol{\mu}^{\textbf{diff}}$ for $\boldsymbol{X}_k$.

Table 4.4 displays the bias, mean squared error (MSE), and coverage rates of $\widehat{\Delta}$ for each data-generating covariate $\boldsymbol{X}_k$ under each scenario, each with 5000 replications. Bolded quantities indicate optimal values; that is, estimated biases within a margin of error from 0, lowest MSEs within

Scenario	Bias	MSE	Coverage	Scenario	Bias	MSE	Coverage
$\boldsymbol{X}_1$				$\boldsymbol{X}_5$			
Choice 1	0.528	0.297	0.038	Choice 1	**−0.003**	0.048	**0.949**
Choice 2	0.500	0.260	0.050	Choice 2	**0.002**	0.045	**0.951**
Choice 3	**0.000**	**0.025**	**0.955**	Choice 3	**0.000**	**0.025**	**0.954**
Choice 4	**0.000**	**0.026**	**0.951**	Choice 4	**0.000**	**0.026**	**0.952**
Choice 5	0.502	0.269	0.054	Choice 5	**0.002**	0.046	**0.956**
Choice 6	0.499	0.267	0.051	Choice 6	**−0.002**	0.046	**0.949**
Choice 7	-0.041	0.028	0.943	Choice 7	**−0.001**	**0.026**	**0.951**
Choice 8	**0.001**	**0.025**	**0.948**	Choice 8	**0.001**	**0.025**	**0.949**
$\boldsymbol{X}_2$				$\boldsymbol{X}_6$			
Choice 1	0.029	**0.013**	0.935	Choice 1	**−0.003**	**0.025**	**0.946**
Choice 2	**0.002**	**0.013**	**0.952**	Choice 2	**0.001**	**0.024**	**0.955**
Choice 3	**0.000**	0.025	**0.955**	Choice 3	**0.000**	**0.025**	**0.955**
Choice 4	**0.000**	0.026	**0.951**	Choice 4	**0.000**	**0.026**	**0.952**
Choice 5	**0.001**	**0.013**	**0.952**	Choice 5	**−0.004**	**0.025**	**0.946**
Choice 6	**−0.001**	**0.013**	**0.953**	Choice 6	**−0.001**	**0.025**	**0.950**
Choice 7	-0.041	0.028	0.943	Choice 7	**−0.001**	**0.026**	**0.951**
Choice 8	**0.001**	0.025	**0.948**	Choice 8	**0.001**	**0.025**	**0.949**
$\boldsymbol{X}_3$				$\boldsymbol{X}_7$			
Choice 1	0.499	0.267	0.049	Choice 1	**0.002**	0.046	**0.953**
Choice 2	0.500	0.267	0.050	Choice 2	**0.002**	0.045	**0.951**
Choice 3	**0.000**	**0.025**	**0.954**	Choice 3	**0.000**	**0.025**	**0.954**
Choice 4	**0.000**	**0.026**	**0.951**	Choice 4	**0.000**	**0.026**	**0.952**
Choice 5	0.502	0.269	0.050	Choice 5	**0.002**	0.046	**0.951**
Choice 6	0.499	0.267	0.051	Choice 6	**−0.002**	0.046	**0.949**
Choice 7	**−0.001**	**0.026**	**0.952**	Choice 7	**−0.001**	**0.026**	**0.951**
Choice 8	**0.001**	**0.025**	**0.948**	Choice 8	**0.001**	**0.025**	**0.949**
$\boldsymbol{X}_4$				$\boldsymbol{X}_8$			
Choice 1	**−0.001**	**0.013**	**0.949**	Choice 1	**−0.003**	**0.025**	**0.947**
Choice 2	**0.002**	**0.013**	**0.952**	Choice 2	**0.001**	**0.024**	**0.955**
Choice 3	**0.000**	0.025	**0.954**	Choice 3	**0.000**	**0.025**	**0.954**
Choice 4	**0.000**	0.026	**0.951**	Choice 4	**0.000**	**0.026**	**0.952**
Choice 5	**0.002**	**0.012**	**0.952**	Choice 5	**0.000**	**0.025**	**0.954**
Choice 6	**−0.001**	**0.013**	**0.953**	Choice 6	**−0.001**	**0.025**	**0.950**
Choice 7	**−0.001**	0.026	**0.952**	Choice 7	**−0.001**	**0.026**	**0.951**
Choice 8	**0.001**	0.025	**0.948**	Choice 8	**0.001**	**0.025**	**0.949**

Table 4.4: Bias, MSE, and Coverage of $\widehat{\Delta}$ under different true data-generating scenarios ($\boldsymbol{X}_k$ panels) and adjustment choices (rows). **Bolded** quantities indicate optimal values.

a margin of error of each other, and coverage rates no more than a margin of error less than 0.95. From these results, we can make the following conclusions:

- Estimation of Δ is biased when we fail to adjust $\boldsymbol{\mu}^{\textbf{diff}}$ for covariates predictive of $\boldsymbol{T}$ and $\boldsymbol{\mu}^{\textbf{diff}}$ (here, Choices 1, 2, 5, and 6 in the $\boldsymbol{X}_1$ and $\boldsymbol{X}_3$ panels).

- Choices 3, 4, and 8 always lead to unbiased estimation of Δ (however, these all require that we observe the covariate $\boldsymbol{X}_k$).

- Estimates of Δ are biased when we do not adjust $\boldsymbol{\mu}$ for covariates predictive of $\boldsymbol{T}$ and $\boldsymbol{\mu}$ (here, Choice 1 in the $\boldsymbol{X}_2$ panel). This bias disappears when we include $\boldsymbol{T}$ as a predictor for $\boldsymbol{\mu}$ (Choices 5 and 6). However, adjusting $\boldsymbol{\mu}^{\textbf{diff}}$ for covariates predictive of $\boldsymbol{\mu}$ and $\boldsymbol{T}$ (here, $\boldsymbol{X}_1$ and $\boldsymbol{X}_3$) without adjusting $\boldsymbol{\mu}$ for those $\boldsymbol{X}_k$ (Choice 7) leads to biased estimates of Δ.

- The most efficient model (by lowest-MSE criterion) varies across covariate cases; in some cases only $\boldsymbol{\mu}$ should be adjusted, in some only $\boldsymbol{\mu}^{\textbf{diff}}$ should be adjusted, in some they should both be adjusted, and in others adjusting either needlessly complicates the model.

- Choices that yield unbiased estimates of Δ have nominal coverage rates, while those with bias do not.

In Table 4.4, when there are covariates predictive of both $\boldsymbol{\mu}$ and $\boldsymbol{T}$ (here, $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$), the bias of $\widehat{\Delta}$ is decreased when $\boldsymbol{T}$ is included as a predictor for $\boldsymbol{\mu}$ (Scenario 1, Scenario 5). However, when such a covariate is excluded from $\boldsymbol{\mu}$, $\boldsymbol{T}$ is retained as a predictor for μ, and the covariate is included for $\boldsymbol{\mu}^{\textbf{diff}}$ (Scenario 7), there is noticeable bias. To avoid this bias, one could design a variable selection technique to force the baseline and change models to include the same predictors. Another variable selection approach to avoid this bias would instead be to include those $\boldsymbol{X}_k$ in the change model that are predictive of $\boldsymbol{T}$ and $\boldsymbol{\mu}^{\textbf{diff}}$, and then only include those $\boldsymbol{X}_k$ in the baseline model that are predictive of $\boldsymbol{T}$, $\boldsymbol{\mu}^{\textbf{diff}}$, and $\boldsymbol{\mu}$.

Variable selection should also be concerned with efficient estimation. When there are covariates predictive of $\boldsymbol{\mu}$ but not $\boldsymbol{\mu}^{\textbf{diff}}$ (here, $\boldsymbol{X}_2$ and $\boldsymbol{X}_6$), bias and MSE are minimized when $\boldsymbol{X}_k$ is included

in the model for $\boldsymbol{\mu}$ but not $\boldsymbol{\mu}^{\textbf{diff}}$ (Scenarios 2 and 6). When there are covariates predictive of $\boldsymbol{\mu}^{\textbf{diff}}$ but not $\boldsymbol{\mu}$ (here, $\boldsymbol{X}_3$ and $\boldsymbol{X}_7$), MSE is minimized when $\boldsymbol{X}_k$ is included in the model for $\boldsymbol{\mu}^{\textbf{diff}}$ regardless of the baseline model adjustment (Scenarios 3, 4, 7, and 8). One approach to variable selection, then, is to perform variable selection on the baseline and change models separately. To arrive at a more parsimonious (i.e., fewer covariates used) model, another variable selection approach is to include any $\boldsymbol{X}_k$ predictive of $\boldsymbol{\mu}^{\textbf{diff}}$ in the change model, and only include any $\boldsymbol{X}_k$ predictive of $\boldsymbol{\mu}$ *and* $\boldsymbol{\mu}^{\textbf{diff}}$ in the baseline model.

Motivated by the application in Chapter 3 and the results of Section 4.2, we now propose four Bayesian variable selection algorithms in the HDiD context.

4.3 Variable Selection Approaches

An often used technique in Bayesian regression is *spike-and-slab* variable selection. Mitchell and Beauchamp[49] first proposed placing priors on each β_k which take the form of a mixture of a point mass at 0 (a "spike") and a uniform distribution over the support of β_k (a "slab"). The idea behind this formulation is that β_k close to 0 will have higher posterior probability under the "spike" and β_k far from 0 will have higher posterior probability under the "slab", and the posterior probability of belonging to each can be used to inform which variables are included or excluded from the final model choice.

This method was extended by George & McCulloch to instead represent the spike-and-slab as a mixture of Gaussian distributions with mean 0, the spike having low variance and the slab having high variance, and presented how to perform this variable selection as a Gibbs sampling algorithm named Stochastic Search Variable Selection.[50] This implementation is formally represented as

$$\begin{aligned} &\beta_k | w_k, z_k^2, c^2 \sim (1 - w_k) N(0, z_k^2) + w_k N(0, c^2 z_k^2), \\ &w_k | p \sim \text{Bernoulli}(p), \ k = 1, ..., K, \end{aligned} \tag{4.4}$$

where $w_k = 1$ indicates β_k was selected for the slab (i.e., $\boldsymbol{X}_k$ is included the model) and $w_k =$

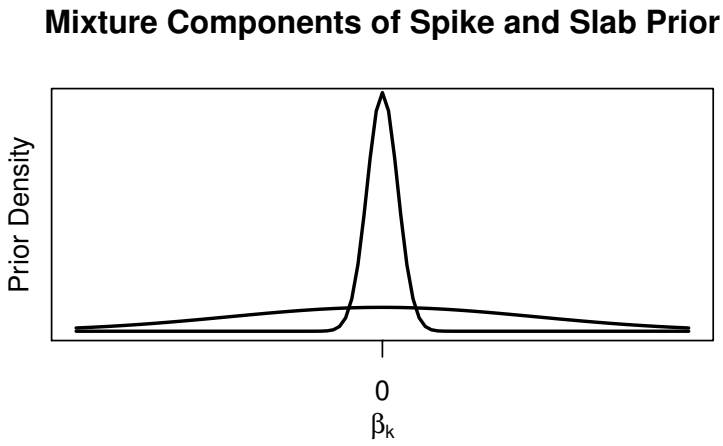


Figure 4.2: Spike-and-slab prior as represented by a mixture of mean zero Gaussian distributions; the curve highly peaked at 0 is the spike component and the more diffuse curve is the slab component.

0 indicates β_k was selected for the spike (i.e., $\boldsymbol{X}_k$ is excluded from the model). To make this formulation work, z_k^2 is chosen to be small and c^2 is chosen to be large. Figure 4.2 illustrates the spike-and-slab prior in Equation (4.4), in which the curve highly peaked at 0 represents the spike and the more diffuse curve represents the slab.

The slab component need not be Gaussian. A central t-distribution works well, as it has good support for values of β_k moderately far from 0 and heavier tails to support values of β_k very far from 0. To implement this, we modify Equation (4.4) to

$$\begin{aligned} &\beta_k | w_k, z_k^2, \lambda_k \sim (1 - w_k) N(0, z_k^2) + w_k t_\nu(0, \lambda_k), \\ &w_k | p \sim \text{Bernoulli}(p), \ k = 1, ..., K, \end{aligned} \tag{4.5}$$

where $t_\nu(0, \lambda_k)$ denotes the central Student t-distribution with ν degrees of freedom and scale λ_k. To facilitate computation, we can take advantage of the t-distribution's equivalence with a scale

mixture of Gaussian distributions, re-expressing Equation (4.5) as

$$\begin{aligned}
&\beta_k|w_k, z_k^2, \gamma_k \sim (1-w_k)N(0, z_k^2) + w_k N(0, 1/\gamma_k), \\
&\gamma_k|\lambda_k \sim \text{Gamma(shape = } \nu/2, \text{rate = } (\nu/2)\lambda_k^2) \\
&w_k|p \sim \text{Bernoulli}(p),\ k = 1, ..., K,
\end{aligned} \tag{4.6}$$

with a diffuse mean-zero Gaussian prior for the intercepts $\widetilde{\beta}_0$ and β_0: $\pi(\widetilde{\beta}_0) = N(0, \widetilde{\omega}^2)$ and $\pi(\beta_0) = N(0, \omega^2)$, where $\widetilde{\omega}^2$ and ω^2 are set to large constants. We now propose four applications of the spike-and-slab prior in Equation (4.5) to perform variable selection in the HDiD context.

4.3.1 Separate Method

The first approach we consider, called the Separate method, performs variable selection separately for the baseline and change models. That is, the posterior of $\widetilde{\beta}_k$ informs whether or not to include $\boldsymbol{X}_k$ in the model for $\boldsymbol{\mu}$, and the posterior of β_k independently informs whether or not to include $\boldsymbol{X}_k$ in the model for $\boldsymbol{\mu}^{\textbf{diff}}$. This approach makes sense when there are some candidate variables related to either the baseline or change models, but perhaps not both. Using this method, the priors for $\widetilde{\beta}_k$ and β_k are

$$\begin{aligned}
&\widetilde{\beta}_k|\widetilde{w}_k, \widetilde{\lambda}_k \sim (1-\widetilde{w}_k)N(0, z_k^2) + \widetilde{w}_k t_\nu(0, \widetilde{\lambda}_k), k = 1, ..., K \\
&\beta_k|w_k, \lambda_k \sim (1-w_k)N(0, z_k^2) + w_k t_\nu(0, \lambda_k), k = 1, ..., K
\end{aligned} \tag{4.7}$$

where $\widetilde{w}_k$ is 1 if $\boldsymbol{X}_k$ is included in the baseline model and 0 otherwise, and w_k is 1 if $\boldsymbol{X}_k$ is included in the change model and 0 otherwise, and $\widetilde{w}_k$ and w_k are assumed to be independent a priori: $\pi(\widetilde{w}_k, w_k) = \pi(\widetilde{w}_k)\pi(w_k)$. We can re-express (4.7) as a multivariate Gaussian distribution using the scale mixture definition of the t-distribution in (4.6). Specifically, we define $a_k = 1/\sqrt{\gamma_k}$ when $w_k = 1$ and $a_k = z_k$ if $w_k = 0$. Then, the prior for $\boldsymbol{\beta}$ arises as $\boldsymbol{\beta} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{D}^2)$, where $\boldsymbol{D} = \text{diag}(\omega, a_1, ..., a_K)$. The setup for $\widetilde{\boldsymbol{\beta}}$ is analogous: $\widetilde{\boldsymbol{\beta}} \sim \boldsymbol{N}(\boldsymbol{0}, \widetilde{\boldsymbol{D}}^2)$.

4.3.2 Shared Method

While the Separate method makes sense when variables are clearly related to either the baseline model or the change model (but not both), there may be scenarios in which some covariates are predictive of both. For example, consider a county-level initiative to encourage its residents to recycle more. If the average socioeconomic status of the county were positively related to baseline per-capita pounds recycled, one would want to include the county's socioeconomic status as a covariate in the baseline model. If residents with higher socioeconomic status were also more willing to change their recycling habits, one would also want to include socioeconomic status in the change model. In this instance, the Separate method would use the posterior masses of $\widetilde{\beta}_k$ and β_k separately to make independent draws for $\widetilde{w}_k$ and w_k, where the optimal approach in this instance would be to use information from both posteriors to strengthen the probability of including a covariate that should be included in both models.

The second approach, which we call the Shared method, is to constrain $\boldsymbol{X}_k$ to be either excluded or included in both models based on the joint posterior distributions of $\widetilde{\beta}_k$ and β_k. We specify a joint prior for $[\widetilde{\beta}_k\ \beta_k]^T$ with a shared inclusion indicator w_k:

$$[\widetilde{\beta}_k\ \beta_k]^T \mid w_k, \lambda_k \sim (1 - w_k)\boldsymbol{N_2}(\boldsymbol{0}, z_k^2\boldsymbol{I}_2) + w_k\boldsymbol{t}_\nu(\boldsymbol{0}, \lambda_k\boldsymbol{I}_2), k = 1, ...K \tag{4.8}$$

where $\boldsymbol{t}_\nu$ is a bivariate t-distribution with ν degrees of freedom.

4.3.3 Sufficient Method

The Separate and Shared methods provide approaches which include $\boldsymbol{X}_k$ if there is statistical evidence that $\boldsymbol{X}_k$ reduces the residual variance in $\boldsymbol{\mu}$ and/or $\boldsymbol{\mu}^{\textbf{diff}}$. These approaches are especially sensible when the number of groups J is small, wherein inference on Δ heavily depends on which covariates are included in the models. Suppose instead that the number of groups J is large, so that estimates of Δ are relatively precise. In this scenario, the only covariates that need to be included to estimate Δ without bias are those related to both $\boldsymbol{T}$ and $\boldsymbol{\mu}^{\textbf{diff}}$. For example, Table 4.4 suggests that

when we adjust the baseline model for $\boldsymbol{T}$ (Choices 5-8), the minimally sufficient set of covariates needed to estimate Δ without bias are those that predict both $\boldsymbol{\mu}^{\textbf{diff}}$ and $\boldsymbol{T}$ (here, $\{\boldsymbol{X}_1, \boldsymbol{X}_3\}$). We propose a third approach, called the Sufficient method, to identify the smallest model that estimates Δ consistently.

First, we introduce an exposure-confounder model to identify those $\boldsymbol{X}_k$ that are associated with $\boldsymbol{T}$. If $\boldsymbol{T}$ is binary, this may be a probit or logistic model. If instead $\boldsymbol{T}$ is continuous, a reasonable exposure model may be

$$\boldsymbol{T} = \boldsymbol{X}\boldsymbol{\alpha} + \boldsymbol{\epsilon}_\alpha, \boldsymbol{\epsilon}_\alpha \sim \boldsymbol{N}(\boldsymbol{0}, \sigma_\alpha^2 \boldsymbol{I}) \tag{4.9}$$

with a noninformative prior on σ_α^2: $\pi(\sigma_\alpha^2) \propto 1/\sigma_\alpha^2$. Then, we can impose a similar spike-and-slab prior on each α_k:

$$\alpha_k | w_k^e, \lambda_k^e \sim (1 - w_k^e) N(0, z_k^e) + w_k^e t_\nu(0, \lambda_k^e)$$

with $\pi(w_k^e) = \text{Bern}(p^e)$, where $w_k^e = 1$ when $\boldsymbol{X}_k$ is selected for this exposure model and is 0 otherwise.

To only include those covariates for $\boldsymbol{\mu}^{\textbf{diff}}$ that are also predictive of $\boldsymbol{T}$, we set $w_k = 0$ if $w_k^e = 0$; that is, $\pi(w_k) = w_k^e * \text{Bern}(p)$. Similarly, to only include those covariates for $\boldsymbol{\mu}$ that are also predictive of $\boldsymbol{\mu}$, $\boldsymbol{\mu}^{\textbf{diff}}$, and $\boldsymbol{T}$, we only allow $\widetilde{w}_k$ to be 1 when $w_k = 1$: $\pi(\widetilde{w}_k) = w_k * \text{Bern}(\widetilde{p})$.

Given the exposure model in (4.9) and outcome model in (3.1), estimation of the treatment effect usually proceeds by first estimating $\boldsymbol{\alpha}$ using the exposure model, treating estimates of $\boldsymbol{\alpha}$ as fixed and known, and then estimating Δ using the outcome model. An alternative fully multivariate approach is to combine (4.9) and (3.1) into one joint likelihood and model both simultaneously in a Bayesian framework. However, Zigler et al. show that such an approach can lead to "feedback" between the two models; that is, quantities in the outcome model informing quantities in the exposure model.[51] In general, model feedback can lead to biased estimates of the treatment effect. To prevent this, we choose to perform variable selection sequentially across the exposure, change, and baseline models. Specifically, within each Gibbs iteration, we draw w_k^e without conditioning on the baseline or change models, and we draw w_k without conditioning on the baseline model.

4.3.4 Efficient Method

Table 4.4 suggests that the most efficient estimates of Δ, in terms of MSE, occur when the covariate set includes those $\boldsymbol{X}_k$ related to $\boldsymbol{\mu}^{\textbf{diff}}$ and excludes those $\boldsymbol{X}_k$ that are not. Combining this with the desire for model parsimony, we blend the Separate and Sufficient methods to include only those $\boldsymbol{X}_k$ related to $\boldsymbol{\mu}^{\textbf{diff}}$. To start, we use the same spike-and-slab prior as in Equation (4.7) and do not fit an exposure model. Since the decision to include $\boldsymbol{X}_k$ in the model for $\boldsymbol{\mu}^{\textbf{diff}}$ depends on whether $\boldsymbol{X}_k$ predicts $\boldsymbol{T}$, we assign a prior for w_k that is independent of the exposure model, say $\pi(w_k) = \text{Bern}(p_k)$. When selecting covariates for the baseline, it suffices to consider only those that are also predictive of change, so we only allow $\widetilde{w}_k$ to be 1 when $w_k = 1$: $\pi(\widetilde{w}_k) = w_k * \text{Bern}(\widetilde{p}_k)$. Again to prevent model feedback, we draw w_k without conditioning on the model for baseline.

4.4 Simulations to Assess Variable Selection Algorithms

We conducted a simulation study to assess bias, MSE, and coverage rates of Δ when implementing the Separate, Shared, Sufficient, and Efficient methods using the covariate set in Section 4.2.2 and data generation scheme described in Section 4.2.3, with $\boldsymbol{\alpha} = [1, 1, 1, 1, 0, 0, 0, 0]^T$, $\widetilde{\boldsymbol{\beta}} = [1, 1, 0, 0, 1, 1, 0, 0]^T$, $\boldsymbol{\beta} = [1, 0, 1, 0, 1, 0, 1, 0]^T$, $\widetilde{\Delta} = 0$, and $\Delta = 1$.

For the slab, we chose a t-distribution with $\nu = 5$ degrees of freedom and $\lambda_k = 5$, which has regression coefficients moderately far from 0, and has thick tails to support β very far from 0. To allow for conditionally conjugate Gibbs sampling, we re-express the slabs for $\widetilde{\beta}_k$ and β_k in (4.7) as $N(0, 1/\widetilde{\gamma}_k)$ and $N(0, 1/\gamma_k)$, with weakly informative Gamma(shape = $5/2$, rate = $(5/2)*5^2$) priors for each $\widetilde{\gamma}_k$ and γ_k. Finally, we chose a spike variance of $z_k^2 = 0.01^2$, prior inclusion probabilities $\widetilde{p} = p = p_e = 1/2$, and prior variance for the intercept coefficients $\widetilde{\omega}^2 = \omega^2 = 10000$.

In the Shared approach, we again took advantage of the scale mixture of Gaussian distributions to represent the t-distribution by re-expressing the model in (4.8) as, for $k = 1, ..., K$:

$$[\widetilde{\beta}_k \ \beta_k]^T \mid w_k, \gamma_k \sim (1-w_k)\boldsymbol{N_2}(\mathbf{0}, z_k^2 \boldsymbol{I}_2) + w_k \boldsymbol{N_2}(\mathbf{0}, 1/\gamma_k \boldsymbol{I}_2), k = 1, ...K$$

	Bias	MSE	Coverage	# predictors, μ^{diff}	# predictors, μ
Full	-0.002 ($\pm$0.005)	0.029 ($\pm$0.001)	0.948	8	8
Separate	0.025 ($\pm$0.004)	0.018 ($\pm$0.001)	0.949	4.12 ($\pm$0.01)	4.15 ($\pm$0.01)
Shared	0.031 ($\pm$0.004)	0.024 ($\pm$0.001)	0.941	5.78 ($\pm$0.01)	5.78 ($\pm$0.01)
Sufficient	0.126 ($\pm$0.006)	0.059 ($\pm$0.002)	0.876	1.69 ($\pm$0.01)	0.52 ($\pm$0.01)
Efficient	0.022 ($\pm$0.004)	0.017 ($\pm$0.001)	0.937	4.13 ($\pm$0.01)	1.75 ($\pm$0.01)
Null	0.407 ($\pm$0.004)	0.185 ($\pm$0.003)	0.180	0	0

Table 4.5: Bias, MSE, and coverage rates (with Margin of Error) of $\widehat{\Delta}$ using the model with all predictors (Full), each variable selection method, and the model with no predictors (Null), each with their mean number of predictors included, with $J = 50$ groups

	$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$	$\boldsymbol{X}_4$	$\boldsymbol{X}_5$	$\boldsymbol{X}_6$	$\boldsymbol{X}_7$	$\boldsymbol{X}_8$
β_k	1	0	1	0	1	0	1	0
$\widetilde{\beta}_k$	1	1	0	0	1	1	0	0
α_k	1	1	1	1	0	0	0	0
Separate	0.929	0.085	0.934	0.085	0.984	0.062	0.984	0.059
Shared	0.993	0.901	0.876	0.063	1.000	0.969	0.958	0.023
Sufficient	0.695	0.139	0.681	0.141	0.012	0.003	0.012	0.003
Efficient	0.936	0.085	0.932	0.086	0.986	0.064	0.984	0.059

Table 4.6: Inclusion probabilities of $\boldsymbol{X}_1, ..., \boldsymbol{X}_8$ in the model for $\boldsymbol{\mu}^{\textbf{diff}}$ using each method with corresponding true values of $\boldsymbol{\beta}$, $\widetilde{\boldsymbol{\beta}}$, $\boldsymbol{\alpha}$, with $J = 50$ groups

	Bias	MSE	Coverage	# predictors, μ^{diff}	# predictors, μ
Full	0.003 ($\pm$0.003)	0.014 ($\pm$0.001)	0.947	8	8
Separate	0.003 ($\pm$0.002)	0.005 ($\pm$0.000)	0.958	4.18 ($\pm$0.00)	4.17 ($\pm$0.00)
Shared	0.004 ($\pm$0.002)	0.008 ($\pm$0.000)	0.953	6.02 ($\pm$0.00)	6.02 ($\pm$0.00)
Sufficient	0.008 ($\pm$0.003)	0.011 ($\pm$0.001)	0.954	2.14 ($\pm$0.00)	1.04 ($\pm$0.01)
Efficient	0.000 ($\pm$0.002)	0.005 ($\pm$0.000)	0.958	4.18 ($\pm$0.00)	2.20 ($\pm$0.01)
Null	0.410 ($\pm$0.003)	0.177 ($\pm$0.002)	0.017	0	0

Table 4.7: Bias, MSE, and coverage rates (with Margin of Error) of $\widehat{\Delta}$ using the model with all predictors (Full), each variable selection method, and the model with no predictors (Null), each with their mean number of predictors included, with $J = 100$ groups

Tables 4.5-4.8 display, for number of groups $J = 50$ and $J = 100$, the bias, MSE, and coverage rates of Δ, the mean number of predictors included in the change and baseline models, and the inclusion probabilities for each predictor. The results are presented for each variable selection

	$\boldsymbol{X}_1$	$\boldsymbol{X}_2$	$\boldsymbol{X}_3$	$\boldsymbol{X}_4$	$\boldsymbol{X}_5$	$\boldsymbol{X}_6$	$\boldsymbol{X}_7$	$\boldsymbol{X}_8$
β_k	1	0	1	0	1	0	1	0
$\widetilde{\beta}_k$	1	1	0	0	1	1	0	0
α_k	1	1	1	1	0	0	0	0
Separate	0.998	0.051	0.998	0.052	1.000	0.040	1.000	0.041
Shared	1.000	0.995	0.991	0.021	1.000	1.000	1.000	0.009
Sufficient	0.975	0.071	0.970	0.071	0.025	0.001	0.023	0.001
Efficient	0.999	0.052	0.998	0.052	1.000	0.042	1.000	0.041

Table 4.8: Inclusion probabilities of $\boldsymbol{X}_1, ..., \boldsymbol{X}_8$ in the model for $\boldsymbol{\mu}^{\textbf{diff}}$ using each method with corresponding true values of $\boldsymbol{\beta}$, $\widetilde{\boldsymbol{\beta}}$, $\boldsymbol{\alpha}$, with $J = 100$ groups

algorithm with the results using the "Full" model (i.e., with every covariate $\{\boldsymbol{X}_1, ..., \boldsymbol{X}_8\}$ for $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$) and the "Null" model (i.e., not adjusting either model for any covariates). Overall, the Separate and Efficient methods performed the best in terms of bias and MSE. The Sufficient method suffered higher bias and MSE with lower coverage; a post-hoc investigation showed that this was due to the added uncertainty in performing variable selection for the exposure model. The Null model performed poorly, having the highest bias and MSE with unacceptably low coverage. While $\hat{\Delta}$ as estimated by the Full model was unbiased, it had higher MSE than the methods with variable selection in almost every scenario. The Shared method included the most predictors for $\boldsymbol{\mu}^{\textbf{diff}}$, while the Sufficient method yielded the most parsimonious model for $\boldsymbol{\mu}^{\textbf{diff}}$. Tables 4.6 and 4.8 show that the inclusion probabilities behaved as desired using each method. Specifically, the Separate and Efficient methods included covariates that were predictive of $\boldsymbol{\mu}^{\textbf{diff}}$ often, the Shared method included covariates that were predictive $\boldsymbol{\mu}^{\textbf{diff}}$ or $\boldsymbol{\mu}$ often, and the Sufficient method only included those covariates that were predictive of $\boldsymbol{\mu}^{\textbf{diff}}$ and $\boldsymbol{T}$ often. The bias and MSE (with their margins of error) decreased as J increased from 50 to 100, suggesting that each variable selection method leads to consistent estimation of Δ. Similarly, the inclusion probabilities approached either 0 or 1 as J increased, suggesting that the Separate, Sufficient, and Efficient methods are consistent in including the variables that they are designed to select. The Shared method tended to include a variable in both models if it is predictive of either baseline ($\widetilde{\beta}_k$) or change (β_k).

4.5 Application to Primary Care Redesign and Diabetes Data

	Age	Female	IVD	Type 1	Comm.	Tobacco	Wealth	Income	$\boldsymbol{c}^{(2008)}$
A1c									
Separate	0.021	0.031	0.337	0.061	0.027	0.104	0.035	0.063	0.008
Shared	0.001	0.978	0.156	0.004	0.001	0.030	0.003	0.003	0.000
Sufficient	0.003	0.008	0.033	0.041	0.014	0.033	0.007	0.003	0.009
Efficient	0.017	0.027	0.282	0.071	0.030	0.106	0.036	0.041	0.008
LDL									
Separate	0.430	0.128	1.000	0.324	0.033	0.139	0.053	0.119	0.018
Shared	0.529	0.956	1.000	0.337	0.003	0.050	0.001	0.002	0.001
Sufficient	0.013	0.010	0.054	0.244	0.016	0.084	0.007	0.006	0.053
Efficient	0.674	0.043	1.000	0.158	0.033	0.130	0.055	0.098	0.019
SBP									
Separate	0.145	0.044	0.561	0.092	0.037	0.115	0.023	0.046	0.010
Shared	0.066	0.002	0.109	0.012	0.002	0.020	0.001	0.001	0.000
Sufficient	0.017	0.008	0.046	0.053	0.020	0.039	0.006	0.004	0.011
Efficient	0.127	0.039	0.579	0.092	0.038	0.101	0.027	0.044	0.010

Table 4.9: Inclusion probabilities for each candidate variable in the change model, A1c

In this section, we use outcome and demographic patient data from MNCM, neighborhood-level covariates from the ACS, and clinic resources and services survey data from the PPCRS to quantify the causal impact of primary care redesign on mean diabetes outcomes from the year 2008 (pre-treatment) to the year 2017 (post-treatment) on $J = 96$ clinics. Similar to the application in Section 3, we identified one principal component[23] driving the variance in the 2008 and 2017 PPCRS results, so we again define a clinic's "score" as the first principal component of the survey matrix. To measure how a clinic matured in its primary care delivery from 2008 to 2017, we again define our main exposure of interest as the clinic score difference: $\boldsymbol{c}^{\mathbf{diff}} = \boldsymbol{c}^{(2017)} - \boldsymbol{c}^{(2008)}$.

For the A1c, LDL, and SBP outcomes separately, we fit the HDiD model in (3.1) and apply each of the four variable selection techniques introduced in Section 4.3. Each candidate predictor is a change in proportions or means from 2008 to 2017; for example, the "Age" predictor is the mean patient age at a particular clinic in 2017 subtracted by its mean age in 2008. For each method, we use a spike standard deviation of $z_k = 0.025$ for each candidate predictor, $k = 1, ..., K$.

Table 4.9 displays the inclusion probabilities within the change model for each candidate predictor and outcome. Change in percent of Female patients seems to be predictive of baseline A1c

and LDL but not change in A1c or LDL, evidenced by the starkly different inclusion probabilities between the Shared approach vs. the others. Change in percent of patients with IVD is a strong predictor of change in LDL, though must not be a strong predictor of treatment, seen in the differences in inclusion probabilities between the Sufficient and Separate, Shared, and Efficient methods. There does not appear to be much evidence that baseline clinic score $\boldsymbol{c}^{(2008)}$ is predictive of change in A1c, LDL, or SBP.

	Separate	Shared	Sufficient	Efficient	Full
$\boldsymbol{c}^{\text{diff}}$	(-0.02, 0.03)	(-0.02, 0.03)	(-0.02, 0.03)	(-0.02, 0.03)	(-0.01, 0.05)
Age	(-0.06, 0.04)	(-0.05, 0.04)	(-0.05, 0.04)	(-0.05, 0.04)	(-0.24, 0.11)
Female	(-0.05, 0.06)	(-0.16, 0.34)	(-0.05, 0.05)	(-0.06, 0.05)	(-0.24, 0.30)
IVD	(-0.82, 0.04)	(-0.86, 0.04)	(-0.07, 0.04)	(-0.82, 0.04)	(-0.96, 0.01)
Type 1	(-0.05, 0.24)	(-0.05, 0.05)	(-0.05, 0.10)	(-0.05, 0.33)	(-0.18, 1.10)
Comm.	(-0.06, 0.04)	(-0.05, 0.04)	(-0.06, 0.04)	(-0.07, 0.04)	(-0.25, 0.14)
Tobacco	(-0.05, 0.56)	(-0.05, 0.62)	(-0.05, 0.07)	(-0.05, 0.53)	(-0.55, 0.85)
Wealth	(-0.07, 0.04)	(-0.06, 0.04)	(-0.06, 0.04)	(-0.08, 0.04)	(-0.24, 0.12)
Income	(-0.13, 0.03)	(-0.06, 0.03)	(-0.06, 0.03)	(-0.09, 0.03)	(-0.22, 0.06)
$\boldsymbol{c}^{(2008)}$	(-0.01, 0.05)	(-0.01, 0.04)	(-0.01, 0.05)	(-0.01, 0.05)	(-0.02, 0.05)

Table 4.10: 95% credible intervals for $\boldsymbol{\beta}$ using each variable selection method, A1c

	Separate	Shared	Sufficient	Efficient	Full
$\boldsymbol{c}^{\text{diff}}$	**(-0.07, -0.00)**	**(-0.07, -0.01)**	**(-0.09, -0.01)**	**(-0.07, -0.00)**	(-0.06, 0.02)
Age	(-0.03, 0.54)	(-0.01, 0.67)	(-0.04, 0.06)	(-0.02, 0.53)	**(0.25, 0.73)**
Female	(-0.04, 0.51)	(-0.04, 0.71)	(-0.05, 0.05)	(-0.05, 0.11)	**(0.01, 0.69)**
IVD	**(-2.38, -1.04)**	**(-2.43, -1.16)**	(-1.57, 0.04)	**(-2.34, -1.06)**	**(-2.23, -0.93)**
Type 1	(-1.45, 0.05)	(-1.52, 0.05)	(-1.52, 0.04)	(-1.06, 0.06)	(-0.76, 1.03)
Comm.	(-0.05, 0.06)	(-0.05, 0.05)	(-0.05, 0.05)	(-0.05, 0.06)	(-0.04, 0.61)
Tobacco	(-0.05, 1.02)	(-0.05, 1.14)	(-0.05, 1.18)	(-0.06, 0.82)	(-0.04, 1.82)
Wealth	(-0.17, 0.04)	(-0.06, 0.04)	(-0.06, 0.04)	(-0.17, 0.04)	(-0.32, 0.17)
Income	(-0.25, 0.03)	(-0.06, 0.03)	(-0.06, 0.04)	(-0.22, 0.03)	(-0.33, 0.06)
$\boldsymbol{c}^{(2008)}$	(-0.01, 0.06)	(-0.01, 0.06)	(-0.00, 0.09)	(-0.01, 0.06)	**(0.00, 0.10)**

Table 4.11: 95% credible intervals for $\boldsymbol{\beta}$ using each variable selection method, LDL

Tables 4.10-4.12 display, using each variable selection method and fitting the full model, the 95% credible intervals for Δ and $\boldsymbol{\beta}$ with the A1c, LDL, and SBP outcomes (respectively). Overall, the variable selection methods led to similar estimates for $\boldsymbol{c}^{\text{diff}}$, a sensible result given the large number of clinics J and that all methods are designed to give consistent estimates of the treatment effect. All credible intervals for the A1c outcome contain 0. The main exposure of interest, $\boldsymbol{c}^{\text{diff}}$

	Separate	Shared	Sufficient	Efficient	Full
$\boldsymbol{c}^{\mathbf{diff}}$	**(-0.09, -0.02)**	**(-0.09, -0.02)**	**(-0.09, -0.02)**	**(-0.09, -0.02)**	**(-0.11, -0.03)**
Age	(-0.04, 0.35)	(-0.04, 0.34)	(-0.04, 0.06)	(-0.04, 0.34)	**(0.07, 0.59)**
Female	(-0.05, 0.10)	(-0.05, 0.05)	(-0.05, 0.05)	(-0.05, 0.12)	(-0.05, 0.68)
IVD	(-0.04, 1.36)	(-0.04, 0.95)	(-0.04, 0.75)	(-0.04, 1.34)	**(0.04, 1.44)**
Type 1	(-0.50, 0.07)	(-0.05, 0.05)	(-0.24, 0.06)	(-0.50, 0.07)	(-0.96, 0.99)
Comm.	(-0.05, 0.09)	(-0.05, 0.05)	(-0.05, 0.06)	(-0.05, 0.08)	(-0.22, 0.49)
Tobacco	(-0.68, 0.06)	(-0.08, 0.05)	(-0.38, 0.05)	(-0.70, 0.06)	(-1.03, 0.94)
Wealth	(-0.05, 0.06)	(-0.05, 0.05)	(-0.05, 0.05)	(-0.05, 0.06)	(-0.21, 0.31)
Income	(-0.04, 0.13)	(-0.04, 0.06)	(-0.04, 0.06)	(-0.04, 0.12)	(-0.11, 0.30)
$\boldsymbol{c}^{\mathbf{(2008)}}$	(-0.05, 0.02)	(-0.05, 0.02)	(-0.05, 0.02)	(-0.05, 0.02)	(-0.08, 0.02)

Table 4.12: 95% credible intervals for $\boldsymbol{\beta}$ using each variable selection method, SBP

is significantly negatively associated with the $\boldsymbol{\mu}^{\mathbf{diff}}$ for the LDL and SBP outcomes, suggesting that greater strides in PCMH redesign lower diabetes patients' cholesterol and blood pressure. The only other significant dynamic predictor is change in proportion of patients with IVD for the change in LDL outcome, suggesting that clinics who gained more patients with IVD from 2009 to 2017 saw the average LDL of their patients decrease.

4.6 Posterior Consistency

One desirable property of a statistical estimator is *consistency*: as the number of data points used to compute the estimator increases indefinitely, the distribution of the estimates become more and more concentrated around the true parameter value. Formally, $\hat{\theta}_n$ is a weakly consistent estimator of θ_0 if $\hat{\theta}_n$ converges to θ_0 in probability as $n \to \infty$; that is, for all $\epsilon > 0$,

$$\lim_{n\to\infty} P(||\hat{\theta}_n - \theta_0|| > \epsilon) = 0, \tag{4.10}$$

where $||\hat{\theta}_n - \theta_0||$ is the norm of $\hat{\theta}_n - \theta_0$ $\left(\text{e.g., the Euclidean norm } \sqrt{(\hat{\theta}_n - \theta_0)^T(\hat{\theta}_n - \theta_0)}\right)$.
A highly similar idea in the Bayesian paradigm is *posterior consistency*. Given a model $f(\boldsymbol{Y}|\theta)$ for data $\boldsymbol{Y} = \{y_1, ..., y_n\} \sim f_0(\boldsymbol{Y}|\theta_0)$ and prior $\pi(\theta)$, the posterior distribution $p(\theta|\boldsymbol{Y})$ is consistent if, for all $\epsilon > 0$,

$$\lim_{n\to\infty} p(||\theta_n - \theta_0|| > \epsilon \mid \boldsymbol{Y}) = 0. \tag{4.11}$$

As the number of groups J increases, we desire the posterior of Δ estimated by our model to concentrate closely around the true value of Δ. We can check three covariate adjustment cases studied in this work to see if they satisfy posterior consistency: (i) not adjusting $\boldsymbol{\mu}$ for covariates, (ii) adjusting $\boldsymbol{\mu}$ for the full set of covariates, and (iii) performing the four variable selection algorithms for $\boldsymbol{\mu}$, in each case. Throughout this section, we will use $\theta_0 = [\widetilde{\beta}_1,\ \widetilde{\beta}_2,\ \widetilde{\beta}_3,\ \Delta,\ \beta_1,\ \beta_2,\ \beta_3]^T = [1,\ 0.5,\ 0,\ 0,\ 0.5,\ 1,\ 0]^T$.

J	20	50	100	150
Bias($\hat{\Delta}$)	0.0909	0.0909	0.0909	0.0909
Bias($\hat{\beta}_1$)	0.0455	0.0455	0.0455	0.0455

Table 4.13: Bias incurred by misspecified baseline model (i) as number of groups J increases

Case (i) is equivalent to placing point mass at 0 with probability 1 priors on $\widetilde{\beta}_1$, $\widetilde{\beta}_2$, and $\widetilde{\beta}_3$, when in reality $\widetilde{\beta}_1$ and $\widetilde{\beta}_2$ are nonzero. Expressing the estimator of θ_0 as $\widehat{\theta} = [0,\ 0,\ 0,\ \hat{\Delta},\ \hat{\beta}_1\ \hat{\beta}_2,\ \hat{\beta}_3]^T$, it is trivial to establish that constraining $\widetilde{\beta}_1$ and $\widetilde{\beta}_2$ to 0 implies that the Euclidean norm $||\widehat{\theta} - \theta_0|| \geq \sqrt{5}/2$, so that failing to adjust $\boldsymbol{\mu}$ for important predictors does not lead to posterior consistency in estimating θ_0. To evaluate the asymptotic behavior of Bias($\hat{\Delta}$) and Bias($\hat{\beta}_1$) individually, we extend the simulation in Section 4.2.1 to number of groups $J = 20$, 50, 100, and 150, computing the theoretical bias established in Theorem 2, and displaying the average bias over 10,000 replications in Table 4.13. The bias stays constant as J increases, evidence that an HDiD model that omits nonzero predictors at *pre*-intervention leads to posterior inconsistency in the estimation of coefficients relating to change. These results match those found in Table 4.1, save for some simulation error.

Doob established surprisingly general conditions under which a Bayesian model enjoys posterior consistency.[52] A modified statement of Doob's Theorem is

Theorem 3. *Assume $\{P_\theta; \theta \in \Theta\}$ is a measurable family of models, θ is identifiable (that is, $P_\theta = P_{\theta'} \Rightarrow \theta = \theta'$), and $y_1, ..., y_n \overset{iid}{\sim} P_{\theta_0}$. If prior $\pi(\theta) > 0\ \forall\ ||\theta - \theta_0|| < \epsilon$, then the posterior distribution of θ is consistent.*

In other words, as long as the prior for θ has nonzero support in every neighborhood of θ_0,

the posterior distribution of θ concentrates tighter in neighborhoods of θ_0 as the sample size (or, in our case, number of groups J) increases indefinitely. Because case (ii) (that is, adjust $\boldsymbol{\mu}$ for all covariates) has nonzero prior support over $\mathbb{R}^{10}$, Theorem 3 guarantees it has posterior consistency. Similarly, for each coefficient, each variable selection algorithm has nonzero support over $\mathbb{R}^K$ and $\mathbb{R}^1$ (or in the Shared approach, $\mathbb{R}^2$) for $\widetilde{\boldsymbol{\beta}}$, $\boldsymbol{\beta}$, and Δ respectively, so Theorem 3 guarantees posterior consistency for them as well. Tables 4.5 and 4.7 provide empirical evidence that the variable selection approaches enjoy posterior consistency; there, the biases incurred by using each method tend to 0 as the number of groups J increases. We should note that the Shared model yielding consistent posteriors for $\widetilde{\boldsymbol{\beta}}$ and $\boldsymbol{\beta}$ does not imply that the variable selection therein is necessarily consistent; that is, as J increases indefinitely, inclusion probabilities for predictors with nonzero association with the response tend to 1 and for predictors with no association with the response tend to 0. However, Tables 4.6 and 4.8 provide evidence that each variable selection technique behaves as it was designed to, asymptotically. As J increases from 50 to 100, the inclusion probabilities in the change model increase (decrease) for those covariates truly predictive (non-predictive) of change using the Separate and Efficient methods, while the inclusion probabilities in the change model increase for those covariates truly predictive of change *or* baseline (and decrease for those truly non-predictive of either) using the Shared method. As J increases from 50 to 100 using the Sufficient method, the inclusion probabilities in the change model increase for only those covariates which are predictive of both $\boldsymbol{T}$ *and* $\boldsymbol{\mu}^{\mathbf{diff}}$ (here, $\boldsymbol{X}_1$ and $\boldsymbol{X}_3$) and decrease for other covariates.

4.7 Conclusions

In this chapter, we introduced a hierarchical extension of the difference-in-differences model and suggested variable selection methods therein. We showed that estimation of the treatment effect is biased if we do not adjust a baseline model either for the treatment, or for all covariates predictive of both the treatment and baseline. We then showed that in order to estimate the treatment effect without bias, we also need to adjust the change model for all covariates that jointly affect both the

treatment and change in outcome. Covariates that are correlated with the change in means should also be included in the model for change to achieve more precise estimates of the treatment effect.

With these guidelines in place, we presented four Bayesian variable selection techniques that can be implemented in the HDiD framework. Just as in our application in Section 4.5, such methods are useful when subjects cannot be matched from pre- to post-treatment timepoints and when the treatment is administered at the group level. Through simulation, we found that each approach leads to reasonable estimation of the treatment effect, and the results suggest that the approaches are consistent. Our simulations suggested that the Sufficient method leads to the smallest covariate set able to estimate the treatment effect consistently, while the Efficient method estimates the treatment effect consistently and with the lowest variance. We applied these methods and show that as a clinic matures in its transformation as a patient-centered medical home, the average LDL and SBP of its patients decreases. Finally, we verified that our variable selection techniques enjoy posterior consistency.

In this chapter, we did not consider the consequences of adjusting for post-treatment covariates whose values are affected by the treatment (that is, those $\boldsymbol{X}_k$ such that $\boldsymbol{T} \rightarrow \boldsymbol{X}_k \rightarrow \boldsymbol{\mu}^{\mathbf{diff}}$), which is extensively cautioned against in the literature.[27,53] We do not recommend applying our variable selection algorithms using such covariates.

Our current approaches assume a somewhat strict prior correlation structure between the inclusion indicators. Specifically, the Separate method assumes a priori that the inclusion indicators for the baseline and change model are uncorrelated, while the Shared method forces the two to be identical. The Sufficient and Efficient methods induce correlation by forcing a covariate excluded from the change model to be excluded from the baseline model. An approach that allows for a more flexible dependence structure between the baseline and change models is introduced in Chapter 6.

Software

The codes to perform the simulation in Section 4.2.5 and the variable selection algorithms in Section 4.3 are available on Github: https://github.com/JamesNormington/VS_in_HDiD.

Chapter 5

Cross-Sectional PPCRS Analysis

5.1 Introduction

One of the primary goals of the UNITED project is to identify the *specific* PCMH-related services and resources that lead to the most improvement in patient diabetes management. So, instead of measuring a single summary measure of clinic primary care transformation, we extend our focus to multiple interventions corresponding to the individual items on the PPCRS.

Recall from our discussion in Chapter 2 that diabetes mellitus is a chronic condition that negatively impacts several aspects of one's health profile. Primary care delivery plays an important role in a patient's diabetes management, and the primary goal of the UNITED team is to investigate the effect of primary care transformation to a patient-centered medical home (PCMH) on diabetes outcomes.

We identified a specific set of 61 clinical services and resources associated with PCMH certification in the 2017 distribution of the PPCRS, and characterized clinics across Minnesota by the presence or absence of these individual systems of care management. These systems of care management may vary in their impact, depending on the characteristics of a clinic. For example, systems of care management that enhance communication among providers may be less impactful in a small clinic, where there is greater opportunity for informal conversation. Likewise, clinic-based weight

management programs may be more impactful in rural areas, where there is less access to commercial weight loss programs. In this chapter, we aim to measure how PCMH-related systems of care management impact diabetes outcomes across clinic subgroups defined by rurality and the size of the clinic. Specifically, we allocate each clinic to one of the three subgroups: large urban clinics, small urban clinics, and rural clinics. Then, we fit a multi-outcome Bayesian cross-sectional model that allows for within-patient correlations across outcomes and clinic-level random effects.

5.2 Data

Like the applications in Chapters 3 and 4, we combine data from the PPCRS, MN Community Measurement (MNCM), and the American Community Survey (ACS) to explore the relationships between optimal diabetes care and primary care clinic organization. Diabetes outcomes are measured using patient-level MNCM data from 2017 that allows us to observe blood sugar control, blood pressure, cholesterol level and smoking status for nearly all patients with diabetes in the state of Minnesota. We merge this with clinic survey data collected by the research team that contains 61 indicators of systems of care management in place at the clinic. This 2017 survey was administered to all primary care practices participating in MNCM reporting in the previous year; more than 70% of practices participated in the survey. We supplement these data sources with socioeconomic data describing the patients' neighborhoods at the ZIP code level from the ACS. After the data were combined, we observed complete data for 199,373 patients in 394 clinics.

We obtained quality data for patients with diabetes from MNCM, based on 2017 clinic encounters. From this data source we were able to observe four measures of diabetes care for use in our analysis: A1c score, LDL, SBP, and a binary indicator of tobacco use (1 if any use at all, 0 otherwise). Natural log transforms were applied to the A1c and LDL outcomes to reduce right-skewness.

Recall that each of the 129 questions on the PPCRS corresponds to one of five domains corresponding to Bodenheimer and Wagner's Chronic Care Model (CCM): clinical information systems (CIS), decision support (DS), delivery system redesign (DSR), health care organization (HCO), and

self-management support (SMS).[16] In its 2017 distribution, the PPCRS was completed by 396 clinics in Minnesota or clinics in bordering communities that had participated in MNCM reporting. We exclude questions specific to diseases other than diabetes (cardiovascular disease, asthma, and depression) from our analysis, leaving $K = 61$ questions remaining. Of these items, SMS was the most common category with 31 (50.8%) questions, followed by DS with 10 (16.4%), CIS with 8 (13.1%), DSR with 7 (11.5%), and HCO with 3 (4.9%). Two questions (coded *g1* and *g2*) were not part of the original PPCRS written in 2008, and thus have no corresponding CCM domain. PPCRS responses were recorded on an ordinal discrete scale, with some questions allowing two responses ($\{$*No*, *Yes*$\}$) and the remaining questions allowing four responses ($\{$*No*, *Yes but needs improvement*, *Yes and works well*, *Don't know*$\}$). A separate analysis of the data (available from the corresponding author) found that the distinction between *Yes but needs improvement* and *Yes and works well* was negatively correlated with achievement of PCMH status, evidence that certified clinics are more self-critical. For this reason, we code missing, *Don't know*, and *No*, responses as 0 and *Yes but needs improvement* and *Yes and works well* responses as 1, to establish a simple binary measure indicating the reported presence of the system of care management.

Each respondent clinic's rurality was defined by mapping its ZIP code to Rural-Urban Commuting Area (RUCA) codes, and mapping its RUCA code to either "Urban" or "Rural".[54] We then split Urban clinics into "Urban Small" and "Urban Large" using the number of diabetes patients to impute the number of providers at the site. Urban Small clinics are defined as having fewer than 455 diabetes patients, equivalent to fewer than 4 full-time equivalent (FTE) providers; Urban Large clinics are defined as having 4 or more imputed FTE providers. The equivalence was determined by dividing the number of patients by an estimated 130 diabetes patients per FTE provider,[55] and rounding to the nearest unit. Our sample size was not large enough to further subdivide rural clinics by size.

We matched the MNCM data to socioeconomic variables derived from the American Community Survey (ACS) to capture characteristics of the patients' neighborhoods. We used the 2016 five-year average survey results aggregated to the ZIP code summary level, matched to patient ZIP

code to describe the environment in which the patient lives and functions. Like the applications in Chapters 3 and 4, Swaney[14] identified two principal components of socioeconomic status in the patients' ZIP code as a function of education, housing costs, use of the Supplemental Nutritional Assistance Program, household income, and family structure, labeled (i) Income/Education and (ii) Wealth, which we use as covariates. Because race and ethnicity were deliberately excluded, we also include the percentage of residents in the ZIP code who identify as white, non-Hispanic as a covariate.

		Urban Large	Urban Small	Rural	Total
	Number of patients	129,944	24,020	45,409	199,373
	Number of clinics	149	112	133	394
Outcomes	A1c	7.0 (1.7)	7.0 (1.7)	6.9 (1.6)	7.0 (1.7)
	LDL	85 (43)	85 (43)	84 (43)	85 (43)
	SBP	126 (18)	126 (18)	128 (16)	127 (18)
	Non-tobacco	86.0%	85.0%	81.8%	84.9%
Age category	< 40	7.2%	6.9%	5.5%	6.7%
	40-44	5.0%	5.5%	4.2%	4.9%
	45-49	8.2%	8.7%	6.4%	7.9%
	50-54	11.5%	12.0%	9.5%	11.1%
	55-59	16.0%	16.2%	14.5%	15.7%
	60-64	17.6%	17.3%	18.0%	17.6%
	65-69	17.2%	17.1%	19.9%	17.8%
	70-75	17.3%	16.3%	22.0%	18.3%
Sex	Female	46.3%	46.7%	45.6%	46.2%
	Male	53.7%	53.3%	54.4%	53.8%
Insurance type	Commercial	46.5%	47.7%	39.0%	44.9%
	Medicaid	10.4%	10.3%	6.1%	9.4%
	Medicare	35.0%	34.0%	46.5%	37.5%
	Dual	3.8%	3.9%	5.4%	4.2%
	Selfpay/Uninsured	2.5%	2.3%	2.0%	2.4%
	SNBC	1.8%	1.8%	0.9%	1.6%
Disease prevalence	Type 1 diabetes	6.1%	5.2%	4.9%	5.7%
	Ischemic vascular disease	15.3%	14.5%	18.7%	16.0%
	Depression	22.9%	24.4%	24.6%	23.4%
Neighborhood characteristics	Income	0.33 (0.55)	0.47 (0.42)	0.41 (0.47)	0.35 (0.55)
	Wealth	0.54 (0.66)	0.56 (0.69)	0.00 (0.37)	0.42 (0.74)
	White, non-Hispanic	81.0%	86.4%	93.0%	84.7%

Table 5.1: Patient demographics. Continuous measures are summarized with medians and interquartile ranges, categorical measures are summarized with proportions.

The 2017 MNCM data include 323,407 observations from 598 clinics within 50 systems. Of these observations, 229,979 of them came from the 396 clinics that had a PPCRS survey response in 2017. Patient records with missing values for biological sex, diabetes type, insurance type, or ACS factors were excluded (4.4%). Of the remaining patients, records with missing or invalid diabetes outcome measures were also excluded (8.0%). Invalid measures were defined as A1c values outside

the range [3%, 25%]; LDL scores greater than 1000 mg/dL; SBP scores less than 50 mmHg less than the diastolic blood pressure reading. Finally, of the remaining patients, records we were unable to match to ACS socioeconomic variables were also excluded (1.4%). In total, 29,736 patient records were excluded due to missing or invalid data, bringing the total sample size to $N = 199,373$ from $J = 394$ clinics. Overall, there are 45,409 (22.8%) records from 133 (33.8%) Rural clinics, 24,020 records (12.0%) from 112 (28.4%) Urban Small clinics, and 129,944 (65.2%) records from 149 (37.8%) Urban Large clinics. The largest system has 73 clinics, while 18 are single-clinic systems. The largest clinic saw 4,107 unique diabetes patients in 2017, the smallest saw 22. Log transforms were applied to the A1c and LDL outcomes to reduce right-skewness. Table 5.1 presents summary statistics of our patient sample.

5.3 Methods

5.3.1 Notation

Throughout, let $s = 1, ..., S = 50$ index systems, $j = 1, ..., J_s$ index clinics within system s, and $i = 1, ..., n_{sj}$ index individuals within clinic j within system s. Let clinics belong to clinic subtype $g \in \{1, 2, 3\}$, where $g = 1$ indicates a clinic is Rural, $g = 2$ indicates a clinic is Urban Small, and $g = 3$ indicates a clinic is Urban Large. Finally, let $d = 1, ..., D = 4$ and $\boldsymbol{Y^{(d)}}$ index and denote the diabetes management outcomes. Specifically, let $\boldsymbol{Y^{(1)}}$ be a vector of log(A1c) measurements, $\boldsymbol{Y^{(2)}}$ be a vector of log(LDL) measurements, $\boldsymbol{Y^{(3)}}$ be a vector of SBP measurements, and $\boldsymbol{Y^{(4)}}$ be a vector of binary indicators of non-tobacco status.

5.3.2 The Model

We now aim to measure how each of the $K = 61$ PPCRS items relates with our four measures of diabetes management. First, the clinic-level binarized PPCRS items were matched to patient-level identifiers collected in a patient-level matrix $\boldsymbol{P}$ that identifies which systems of care management the patient had available.

Because a patient's diabetes outcomes also depend on factors unrelated to the systems of care management measured by the PPCRS, we adjust our model for characteristics of the clinic, the patient, and the patient's residential neighborhood. Clinic-level characteristics include its system's size and an indicator of whether or not the clinic is a federally qualified health center (FQHC). System size is classified as "large" if the clinic is part of a medical group owning 12 or more primary care clinics, "medium" if the medical group owns 2-11 clinics, and "small" if the clinic is a single site. The "large" clinic size was used as the reference category. Patient-level characteristics include biological sex (= 1 if female, = 0 otherwise), an ischemic vascular disease (IVD) indicator (= 1 if patient has IVD, = 0 otherwise), a depression indicator (= 1 if patient has a depression diagnosis recorded, = 0 otherwise), indicators of insurance type (Medicaid, Medicare, Dual Medicare/Medicaid coverage, self-pay, with Commercial coverage as reference value), age as categorical variable (under 40, then 5-year categories 40-44 through 65-69 and 70-75), and a type 1 diabetes indicator. The patient's neighborhood-level ACS variables included the Income/Education and Wealth components computed using Swaney's method, and percentage in the ZIP code who are White, non-Hispanic. Clinic-level characteristics were matched to the patient- and neighborhood-level characteristics, and all combined in a matrix $\boldsymbol{X}$.

Diabetes management can also be influenced by unobserved personal characteristics such as willpower, resilience, illness perception, and proactive coping.[56] His genetic makeup may also influence control of his A1c, LDL, and SBP levels, as well as playing a role in whether or not he uses tobacco. The effects of these unobserved individual-level traits would then impact A1c, LDL, SBP, and non-tobacco status. To account for these unmeasurable individual effects, we allow a patient's residuals to correlate across outcome models. Finally, we add a clinic-level random effect $\delta_{sj}^{(d)}$ for each clinic j and outcome d to account for unobserved clinic characteristics that are invariant across patients.

Given these considerations, our model is

$$\begin{aligned} Y_{sji}^{(d)} \mid j \in g &= \boldsymbol{P}_{sji}^T \boldsymbol{\beta}_{\boldsymbol{g}}^{(\boldsymbol{d})} + \boldsymbol{X}_{sji}^T \boldsymbol{\beta}_{\boldsymbol{X}}^{(\boldsymbol{d})} + \delta_{sj}^{(d)} + \epsilon_{sji}^{(d)}, \text{ for } d = 1, 2, 3, \text{ and} \\ P(Y_{sji}^{(4)} = 1) \mid j \in g) &= \Phi(\boldsymbol{P}_{sji}^T \boldsymbol{\beta}_{\boldsymbol{g}}^{(\boldsymbol{4})} + \boldsymbol{X}_{sji}^T \boldsymbol{\beta}_{\boldsymbol{X}}^{(\boldsymbol{4})} + \delta_{sj}^{(4)}) \end{aligned} \tag{5.1}$$

where Φ is the cumulative distribution function of the standard normal distribution. The $\delta_{sj}^{(d)}$ are assumed to be independent across outcomes for simplicity: $\delta_{sj}^{(d)} \overset{iid}{\sim} N(0, \sigma_\delta^{2(d)})$. We allow the effects for each question to vary by clinic subtype, but use a hierarchical model that allows for borrowing of information across subtypes by shrinking to a common mean effect; specifically, for each predictor in $\boldsymbol{P}$, $k = 1, ..., K$, we assume $\beta_{1k}^{(d)}, \beta_{2k}^{(d)}, \beta_{3k}^{(d)} \sim N(\bar{\beta}_k^{(d)}, \lambda_d^2)$.

To facilitate a conditionally conjugate Gibbs sampling model, we augment the probit model for the binary non-tobacco outcome as described in Albert and Chib.[57] Specifically, for each $Y_{sji}^{(4)}$, we introduce a latent variable Y_{sji}^* such that

$$Y_{sji}^* \mid j \in g = \boldsymbol{P}_{sji}^T \boldsymbol{\beta}_{\boldsymbol{g}}^{(\boldsymbol{4})} + \boldsymbol{X}_{sji}^T \boldsymbol{\beta}_{\boldsymbol{X}}^{(\boldsymbol{4})} + \delta_{sj}^{(4)} + \epsilon_{sji}^*$$

where $Y_{sji}^{(4)}$ is 1 when $Y_{sji}^* > 0$ and 0 otherwise. By construction, each ϵ_{sji}^* follows a standard normal distribution. The two statements of the probit regression model are equivalent, and the latter will be used for computation purposes. We allow correlation among the residuals by specifying $[\epsilon_{sji}^{(1)}\ \epsilon_{sji}^{(2)}\ \epsilon_{sji}^{(3)}\ \epsilon_{sji}^*]^T \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{\Sigma}_y)$, where we impose no structure (i.e., we do not assume anything about its form) on the covariance matrix $\boldsymbol{\Sigma}_y$ other than $\text{Var}(\epsilon_{sji}^*) = 1$ for $i = 1, ..., n_{sj}$, $j = 1, ..., J$, and $s = 1, ..., S$.

Due to the natural hierarchical structure of the data, we elect to proceed with a fully Bayesian analysis. We place an inverse-Wishart prior on $\boldsymbol{\Sigma}_y$: $\pi(\boldsymbol{\Sigma}_y) \propto W^{-1}(\boldsymbol{I_4}, 4)$. Then, following Cowles, Carlin, and Connett, we constrain the bottom rightmost element of $\boldsymbol{\Sigma}_y$ to be 1.[58] Because they are well-identified by the model and for ease of computation, we place conditionally conjugate non-informative priors on the remaining hyper-parameters: $\pi(\bar{\beta}_k^{(d)}) \propto 1$, $\pi(\sigma_\delta^{2(d)}) = 1/\sigma_\delta^{2(d)}$, $\pi(\lambda_d^2) \propto 1/\lambda_d^2$, and $\pi(\boldsymbol{\beta}_{\boldsymbol{X}}^{(\boldsymbol{d})}) \propto \boldsymbol{1}$, for $k = 1, ..., K = 61$ and $d = 1, ..., D = 4$. We then

perform posterior estimation with conditionally conjugate Gibbs sampling, leveraging the results of Albert and Chib to perform sampling with the binary non-tobacco outcome.[57] See Section B.3 in Appendix B for a full description of the Gibbs sampler.

5.3.3 Criteria for Identifying Positive PPCRS Items

Both positive and negative associations were observed across clinic rurality subtypes and outcomes. Because this research was conducted to guide clinics in making actionable steps to improve health care delivery, we elect to focus only on a subset of PPCRS items that have consistently positive signals across outcomes. Specifically, we identify a PPCRS item as being associated positively with diabetes management if it fulfills two criteria:

- **Criteria 1**: The item is significantly associated with at least two diabetes outcomes in the desirable direction (i.e., decreased A1c, LDL, or SBP; increased P(Non-tobacco)).
- **Criteria 2**: The item is not significantly associated with any outcome in the undesirable direction.

In Criteria 1 and 2, item k within subtype g for outcome d is "significant" if the Bayesian credible interval constructed using the 2.5% and 97.5% percentiles of the posterior for $\beta_{gk}^{(d)}$ does not contain 0. The full set of results are available in Appendix C.

5.4 Results

Table 5.2 displays, for each PPCRS item satisfying Criteria 1 and 2 above, the 95% credible intervals among each clinic rurality subtype. Also displayed alongside each item is its corresponding question domain in Bodenheimer and Wagner's CCM. Among Rural clinics, only item *d20* (use of an interactive website to support patient self-management) was associated with every single outcome in the desirable direction. Overall, items relating to Bodenheimer and Wagner's "self-management support" category were most represented in the Rural subtype, and the most commonly significant

outcomes were LDL, SBP, and non-tobacco status. Among Urban Small clinics, all but one CCM domain contain at least one significantly positive item. The most commonly significant outcomes were A1c and SBP. Every CCM domain is represented approximately equally in the Urban Small panel of Table 5.2. Among Urban Large clinics, only item *c13* (encourage patients to see their personal physician) was significantly associated with every outcome in the desirable direction. LDL was the most commonly significant outcome among the selected items. Urban Large was the clinic rurality subtype with the most patients, which may explain why it yielded the highest number of significant PPCRS items. Overall, only one item satisfied Criteria 1 and 2 across each clinic rurality subtype: *c33* (guideline-based reminders for age-appropriate risk assessments).

Tables 5.3 and 5.4 display the sample correlation matrices of the residuals and raw outcomes. While the directions of the relationships are expected, their magnitudes are surprisingly low; the highest absolute correlation is 0.070 for the outcomes marginally and 0.065 for the residuals. This suggests that our findings would have been similar had the outcomes been modeled independently. Further, it suggests one may lose information when modeling the D5 standard of care as one summarized measure compared to modeling its individual components.

Table 5.5 displays 95% credible intervals for $\boldsymbol{\beta}_{\boldsymbol{X}}^{(d)}$, with the significant intervals color coded by whether or not the direction of the relationship is desirable. Having IVD is significantly related to each outcome: lower LDL, SBP, and probability of non-tobacco status, and higher A1c. Having Medicaid is related with worsened diabetes outcomes, except for significantly lower SBP. Medicare seems to be significantly associated with positive outcomes, except for a higher probability of using tobacco. Patients with self-payment as an insurance type have significantly worse diabetes outcomes across the board. As age category increases, A1c, LDL, and probability of non-tobacco become more significant in their desirable direction, while SBP rises. Notably, all but one $\boldsymbol{\beta}_{\boldsymbol{X}}^{(d)}$ coefficient was significantly related to SBP.

The clinic random effects $\delta_{sj}^{(d)}$ were estimated to be very small. Specifically, the posterior means ranged from -0.00015 to 0.00015 across each outcome. This suggests that the most relevant information at the clinic level is already captured by the PPCRS and other clinic-level covariates.

5.5 Discussion

Among Rural and Urban Small clinics, we identified six PPCRS items each which were significantly associated with desirable diabetes outcomes. Among Urban Large clinics, we identified 13 such PPCRS items. Because the distribution of CCM domains were approximately proportional within each clinic rurality subtype, any mechanisms driving inter-rurality disparities in the relationship between primary care delivery and diabetes outcomes are not made apparent here.

We found that four of the D5's components had low correlations, suggesting that these outcomes are nearly independent. Analyses which summarize these outcomes into a single measure would not measure PPCRS items' association with each individual outcome, relationships which could provide insight into exactly what primary care services lead to optimal diabetes management.

Because the analysis is cross-sectional, the main limitation of the analysis in this chapter is the difficulty in making causal claims. While the significant associations reported here may indeed be causal relationships, we do not claim that our analysis proves these associations are causal. Another limitation is that we only include those clinics which, in 2017, both reported diabetes measures to MNCM and completed the PPCRS. Clinics satisfying these criteria are likely more dedicated to diabetes management than those clinics who do not, so our estimates are susceptible to this selection bias.

Our model assumes a priori that the clinic random effects and regression coefficients on the PPCRS items are independent across outcomes. One could extend the model to instead allow for dependence, but because we observed negligible random effects and little correlation between the residuals, we do not expect that doing so would change the results substantially.

	Service *Does your clinic...*	Domain	log(A1c)	log(LDL)	SBP	Non-tobacco
Rural	*... have a system for tracking laboratory tests until results are available to the clinician?* (b1)	CIS	(-0.003, 0.013)	(-0.012, 0.026)	**(-1.869, -0.520)**	**(0.013, 0.121)**
	... have guideline-based reminders for age-appropriate risk assessments (c33)	DS	(-0.008, 0.015)	**(-0.066, -0.010)**	(-0.173, 1.832)	**(0.112, 0.268)**
	... have a systematic approach to identify and remind patients with chronic illnesses who are due for a follow-up visit? (c23)	SMS	(-0.023, 0.003)	(-0.037, 0.031)	**(-2.939, -0.594)**	**(0.015, 0.182)**
	... routinely provide written materials that explain to the patient the recommended medical care guidelines for their illness? (d17)	SMS	(-0.006, 0.008)	**(-0.040, -0.008)**	**(-1.934, -0.795)**	(-0.042, 0.052)
	... routinely use an interactive web-site sponsored by your organization to support self-management for patients and their families? (d20)	SMS	**(-0.016, -0.004)**	**(-0.048, -0.019)**	**(-1.147, -0.096)**	**(0.007, 0.097)**
	... routinely use and exchange data with patients who have access to their own electronic health record or personal health record to support self-management for patients and their families? (d22)	SMS	(-0.008, 0.005)	**(-0.031, -0.002)**	(-0.307, 0.731)	**(0.015, 0.103)**
Urban Small	*... have medication lists?* (c10)	CIS	(-0.015, 0.001)	(-0.016, 0.024)	**(-1.74, -0.030)**	**(0.063, 0.182)**
	... have guideline-based reminders for age-appropriate risk assessments (c33)	DS	**(-0.023, -0.001)**	(-0.040, 0.014)	**(-2.200, -0.276)**	**(0.023, 0.183)**
	... use its scheduling system to encourage patients to see their personal physician? (c13)	DSR	**(-0.013, -0.001)**	(-0.003, 0.026)	(-0.485, 0.541)	**(0.006, 0.099)**
	... routinely use an interactive web-site sponsored by your organization to support self-management for patients and their families? (d20)	SMS	**(-0.015, -0.000)**	**(-0.050, -0.014)**	(-0.769, 0.555)	(-0.059, 0.053)
	... routinely use and exchange data with patients who have access to their own electronic health record or personal health record to support self-management for patients and their families? (d22)	SMS	(-0.012, 0.000)	**(-0.065, -0.037)**	**(-1.082, -0.102)**	(-0.039, 0.053)
	... have a system to identify patients at high risk for hospitalizations and/or emergency department visits? (g1)	–	**(-0.023, -0.008)**	(-0.017, 0.016)	**(-1.940, -0.761)**	(-0.003, 0.103)
Urban Large	*... have medication lists?* (c10)	CIS	**(-0.012, -0.001)**	**(-0.030, -0.011)**	(-0.045, 0.582)	**(0.034, 0.101)**
	... use checklists of tests or interventions that are needed for prevention or monitoring of chronic illness? (c15)	CIS	(-0.012, 0.001)	**(-0.068, -0.041)**	**(-1.686, -0.699)**	(-0.054, 0.045)
	... maintain a registry of patients with diabetes? (c1)	DS	(-0.008, 0.001)	**(-0.021, -0.003)**	**(-0.888, -0.264)**	(-0.025, 0.041)
	... provide guideline-based reminders for services the patient should receive that appear when seeing the patient for diabetes? (c19)	DS	(-0.007, 0.000)	**(-0.019, -0.005)**	(-0.305, 0.206)	**(0.019, 0.073)**
	... have guideline-based reminders for age-appropriate risk assessments (c33)	DS	**(-0.038, -0.012)**	(-0.007, 0.049)	(-0.944, 1.024)	**(0.087, 0.250)**
	... use its scheduling system to encourage patients to see their personal physician? (c13)	DSR	**(-0.010, -0.002)**	**(-0.015, -0.000)**	**(-0.650, -0.128)**	**(0.010, 0.065)**
	... have a formal process for measuring performance for individual physicians or for the practice site? (f1)	HCO	(-0.009, 0.001)	(-0.017, 0.005)	**(-1.869, -1.101)**	**(0.032, 0.115)**
	... have a systematic approach to identify and remind patients with chronic illnesses who are due for a prescription renewal? (c25)	SMS	**(-0.024, -0.005)**	**(-0.048, -0.013)**	(-0.492, 0.656)	(-0.046, 0.068)
	... have a system to identify and notify patients who are due for age-appropriate clinical preventive services? (c35)	SMS	(-0.003, 0.006)	**(-0.022, -0.002)**	**(-1.924, -1.222)**	(-0.046, 0.027)
	... have a systematic process to screen or assess patients for dementia (age 75+)? (d5)	SMS	(-0.007, 0.011)	**(-0.076, -0.035)**	(-0.495, 0.972)	**(0.006, 0.147)**
	... provide or refer patients to formal support programs to assist in self-management for substance abuse? (d7)	SMS	(-0.008, 0.002)	**(-0.026, -0.007)**	**(-1.304, -0.639)**	**(0.004, 0.072)**
	(d17)	SMS	(-0.005, 0.002)	**(-0.031, -0.016)**	**(-1.976, -1.465)**	(-0.001, 0.051)
	... have a system to manage such high risk patients to try to prevent unnecessary hospitalizations or visit emergency departments? (g2)	–	(-0.016, 0.002)	**(-0.063, -0.022)**	**(-2.352, -0.930)**	(-0.030, 0.099)

Table 5.2: 95% credible intervals for PPCRS items most predictive of optimal diabetes outcomes by Criteria 1 and 2 in Section 5.3.3, with intervals not containing 0 highlighted in **green**. Question domains correspond to the domains of Bodenheimer and Wagner's Chronic Care Model: health care organization (HCO), delivery system redesign (DSR), clinical information systems (CIS), decision support (DS), self-management support (SMS), or none (–).

	log(A1c)	log(LDL)	SBP	Non-tobacco
log(A1c)	1			
log(LDL)	0.044	1		
SBP	0.065	0.062	1	
Non-tobacco	-0.006	-0.009	-0.009	1

Table 5.3: Estimated correlation of the residuals

	log(A1c)	log(LDL)	SBP	Non-tobacco
log(A1c)	1			
log(LDL)	0.070	1		
SBP	0.050	0.041	1	
Non-tobacco	-0.041	-0.030	-0.005	1

Table 5.4: Estimated correlation of the raw outcomes

	log(A1c)	log(LDL)	SBP	Non-tobacco
Female	**(-0.020, -0.001)**	**(0.080, 0.098)**	**(-1.811, -1.771)**	**(0.150, 0.180)**
IVD	**(0.005, 0.030)**	**(-0.120, -0.095)**	**(-1.005, -0.970)**	**(-0.151, -0.111)**
Depression	(-0.010, 0.012)	(-0.011, 0.011)	**(-0.520, -0.482)**	**(-0.224, -0.192)**
% WNH	(-0.001, 0.000)	(-0.001, 0.000)	**(0.013, 0.025)**	**(-0.005, -0.003)**
Wealth	(-0.012, 0.014)	(-0.019, 0.008)	(-0.205, 0.072)	**(0.137, 0.185)**
Income	(-0.029, 0.005)	(-0.007, 0.029)	**(-0.937, -0.562)**	**(0.030, 0.097)**
MSHO	(-0.008, 0.040)	(-0.036, 0.013)	**(-0.968, -0.847)**	**(-0.587, -0.520)**
Medicaid	**(0.007, 0.040)**	**(0.018, 0.052)**	**(-0.147, -0.042)**	**(-0.389, -0.340)**
Medicare	(-0.024, 0.003)	**(-0.046, -0.018)**	**(-0.939, -0.889)**	**(-0.397, -0.355)**
SNBC	(-0.032, 0.040)	(-0.053, 0.021)	**(-1.356, -1.187)**	**(-0.719, -0.624)**
Self-pay	**(0.015, 0.078)**	**(0.036, 0.100)**	**(0.110, 0.351)**	**(-0.262, -0.168)**
Age 40-44	(-0.041, 0.014)	**(-0.066, -0.010)**	**(1.757, 1.962)**	(-0.049, 0.031)
Age 45-49	(-0.038, 0.010)	**(-0.081, -0.031)**	**(2.504, 2.707)**	**(0.016, 0.090)**
Age 50-54	(-0.044, 0.001)	**(-0.089, -0.041)**	**(3.601, 3.807)**	**(0.023, 0.091)**
Age 55-59	**(-0.060, -0.017)**	**(-0.115, -0.070)**	**(4.263, 4.470)**	**(0.086, 0.152)**
Age 60-64	**(-0.077, -0.034)**	**(-0.145, -0.101)**	**(5.275, 5.485)**	**(0.265, 0.330)**
Age 65-69	**(-0.087, -0.041)**	**(-0.157, -0.109)**	**(6.154, 6.367)**	**(0.636, 0.709)**
Age 70-75	**(-0.100, -0.051)**	**(-0.184, -0.134)**	**(6.333, 6.543)**	**(0.846, 0.923)**
Type 1	**(0.062, 0.102)**	(-0.017, 0.025)	**(-1.945, -1.837)**	**(0.010, 0.073)**
FQHC	**(0.015, 0.083)**	**(0.031, 0.106)**	**(3.876, 4.674)**	(-0.066, 0.059)

Table 5.5: 95% credible intervals of $\beta_X^{(d)}$. Desirably significant relationships are marked **green**, undesirably significant relationships are marked **red**

Chapter 6

Variable Selection with Structured Priors

Our variable selection algorithms in Chapter 4 assume a strict prior correlation structure between the inclusion indicators. The Separate method assumes a priori that the inclusion indicators for the baseline and change model are uncorrelated, while the Shared method forces the two to be identical. The Sufficient and Efficient methods induce correlation by forcing a covariate excluded from the change model to be excluded from the baseline model. These assumptions may be unreasonable in some contexts, and instead an approach which allows for a more flexible dependence structure via its prior may be favorable. In this chapter, we develop a variable selection algorithm which leverages the temporal and hierarchical structure in the HDiD context to inform variable selection.

Imposing structure on the inclusion indicators $\widetilde{w}_k$ and w_k in a Bayesian variable selection framework is a topic not well-explored in the statistical literature. Murray, Hobbs, and Carlin propose some structure on these indicators, placing iid Bernoulli(p) priors on the inclusion indicators and a Beta hyperprior on p in a modified commensurate prior context.[59] This method leverages information in external data sets to inform multiple model parameters.

6.1 Structured Method

We now introduce another variable selection approach we call the "Structured" method which enjoys both the robustness of the Separate method and the information use of the Shared method while avoiding the drawbacks of each. We first assume that $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$ are associated with group-level predictors collected in a matrix $\boldsymbol{P}$. In this context, we suppose $\boldsymbol{P}$ has a natural structure in that each variable correlates with a set of similar variables (e.g., the PPCRS' question domains corresponding to Bodenheimer and Wagner's CCM model). We call these groups "domains", indexed by $q = 1, ..., Q$. We apply spike-and-slab priors to each $\widetilde{\Delta}_{qz}$ and Δ_{qz}, $q = 1, ..., Q$, $z = 1, ..., m_q$:

$$
\begin{aligned}
\boldsymbol{\mu} &\sim \boldsymbol{N_J}(\boldsymbol{P}\widetilde{\boldsymbol{\Delta}}, \widetilde{\tau}^2 \boldsymbol{I_J}) \\
\boldsymbol{\mu}^{\textbf{diff}} &\sim \boldsymbol{N_J}(\boldsymbol{P}\boldsymbol{\Delta}, \tau^2 \boldsymbol{I_J}) \\
\widetilde{\Delta}_{qz} &\sim (1 - \widetilde{v}_{qz}) N(0, \widetilde{c}_{qz}^2) + \widetilde{v}_{qz} N(0, \widetilde{\omega}_{qz}^2) \\
\Delta_{qz} &\sim (1 - v_{qz}) N(0, c_{qz}^2) + v_{qz} N(0, \omega_{qz}^2)
\end{aligned}
\tag{6.1}
$$

where $v_{qz} = 1$ ($\widetilde{v}_{qz} = 1$) when $\boldsymbol{P}_{qz}$ is included in the model for $\boldsymbol{\mu}^{\textbf{diff}}$ ($\boldsymbol{\mu}$) and $v_{qz} = 0$ ($\widetilde{v}_{qz} = 0$) when $\boldsymbol{P}_{qz}$ is excluded from the model for $\boldsymbol{\mu}^{\textbf{diff}}$ ($\boldsymbol{\mu}$).

To allow for items within the same domain to correlate, our priors for $\widetilde{v}_{qz}$ and v_{qz} should include a domain q-specific effect. We also want to relax the strict prior assumptions on the correlation between inclusion indicators $\widetilde{\boldsymbol{v}}$ and $\boldsymbol{v}$ imposed by the Separate and Shared methods. To flexibly model the correlation between $\widetilde{v}_{qz}$ and v_{qz}, we again leverage the latent variable representation of the probit model, explicitly modeling the correlation between item-qz specific effects across baseline and change.

With these considerations in mind, a reasonable set of priors for the inclusion indicators is:

$$
\begin{aligned}
&\widetilde{v}^*_{qz} = \widetilde{\nu}_q + \widetilde{\epsilon}_{qz}, \\
&v^*_{qz} = \nu_q + \epsilon_{qz}, \\
&\begin{bmatrix} \widetilde{\epsilon}_{qz} \\ \epsilon_{qz} \end{bmatrix} \sim \boldsymbol{N_2}(\mathbf{0_2}, \boldsymbol{\Omega}), \text{ with } \boldsymbol{\Omega} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \\
&\pi(\widetilde{\nu}_q) = \pi(\nu_q) = N(0,1) \\
&\pi(\rho) = (1-\rho^2)^{\eta-1}
\end{aligned}
\tag{6.2}
$$

where $\widetilde{v}^*_{qz} > 0$ when $\widetilde{v}_{qz} = 1$ and $\widetilde{v}^*_{qz} < 0$ when $\widetilde{v}_{qz} = 0$, $v^*_{qz} > 0$ when $v_{qz} = 1$ and $v^*_{qz} < 0$ when $v_{qz} = 0$, and the prior on ρ comes from Lewandowski, Kurowicka, and Joe (the "LKJ" prior).[60]

6.2 Simulation

Here, we describe two simulations to compare the Structured method against the Separate and Shared methods. In each simulation, we identify five domains indexed by $q = 1, ..., Q = 5$, each domain having $m_q = 20$ items. Then, we specify $\widetilde{\boldsymbol{\Delta}}$ and $\boldsymbol{\Delta}$ to control which domains differ in how strongly predictive of $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\textbf{diff}}$ their items are. We go on to generate data by the following scheme:

1. $\boldsymbol{\mu} \sim \boldsymbol{N_J}(\boldsymbol{P}\widetilde{\boldsymbol{\Delta}}, \boldsymbol{I_J})$
2. $\boldsymbol{\mu}^{\textbf{diff}} \sim \boldsymbol{N_J}(\boldsymbol{P}\boldsymbol{\Delta}, \boldsymbol{I_J})$
3. For groups $j = 1, ..., J = 100$, generate $Y^{(0)}_{ji} \sim N(\mu_j, 1)$ and $Y^{(1)}_{ji} \sim N(\mu_j + \mu^{\text{diff}}_j, 1)$ for $i = 1, ..., 10$.

For each simulation, we fix a spike standard deviation of $\widetilde{c}_{qz} = c_{qz} = 0.1$ and a slab standard deviation of $\widetilde{\omega}_{qz} = \omega_{qz} = 5$ for each item z within each domain q, and we fix $\eta = 1$ to impose a $U(-1, 1)$ prior on ρ. We use 5,000 replications, applying the Structured, Separate, and Shared

methods. Within each replication, we perform a combination of conditionally conjugate Gibbs sampling and Metropolis Hastings steps for estimation. The full MCMC algorithm is presented in Section B.4 of Appendix B.

In Simulation 1, we want to compare the Structured method with the Separate and Shared methods when there is (i) domain structure and (ii) perfect correlation between $\widetilde{\mathbf{\Delta}}$ and $\mathbf{\Delta}$. To achieve this, in the first domain ("Domain 1"), each item is predictive of both baseline and change (here, $\widetilde{\Delta}_{1z} = \Delta_{1z} = 1$ for each $z = 1, ..., m_z = 20$). Then, 15 items in Domain 2 are predictive while 5 are not, with each $\widetilde{\Delta}_{2z} = \Delta_{2z}$. We similarly define Domain 3 with 10 predictive items and 10 non-predictive items Domain 4 with 5 predictive and 15 non-predictive items, and Domain 5 with 20 non-predictive items.

In Simulation 2, we want to compare the methods when there is (a) no domain structure and (b) perfect correlation between $\widetilde{\mathbf{\Delta}}$ and $\mathbf{\Delta}$. To achieve this, we again set $\widetilde{\mathbf{\Delta}} = \mathbf{\Delta}$, but specify 10 predictive and 10 non-predictive items in each domain.

Results

	Bias			**RMSE**			**Inc. Prob.**		
Domain	Str.	Sep.	Shr.	Str.	Sep.	Shr.	Str.	Sep.	Shr.
1 (all 1's)	-0.007	-0.118	-0.062	0.193	0.532	0.388	0.991	0.770	0.896
2 (1's)	-0.007	-0.117	-0.066	0.193	0.534	0.386	0.990	0.770	0.895
0's	-0.000	0.007	-0.002	0.087	0.292	0.231	0.035	0.189	0.073
3 (1's)	-0.012	-0.116	-0.063	0.198	0.532	0.388	0.987	0.770	0.898
0's	0.001	0.001	0.002	0.069	0.277	0.219	0.018	0.182	0.070
4 (1's)	-0.013	-0.125	-0.074	0.197	0.524	0.391	0.984	0.773	0.890
0's	-0.001	-0.004	-0.003	0.072	0.291	0.218	0.011	0.190	0.070
5 (all 0's)	0.000	0.002	-0.000	0.069	0.286	0.215	0.005	0.186	0.067

Table 6.1: Average bias, root mean squared error (RMSE), and inclusion probabilities (Inc. Prob.) for predictors within each Domain, split by those nonzero predictors (1's) and zero predictors (0's), comparing the Structured (Str.), Separate (Sep.), and Shared (Shr.) approaches with heterogenous domains

Tables 6.1 and 6.2 display, for each domain, the mean bias, root mean squared error (RMSE), and average inclusion probabilities, across the items within the domain separately for those items

	Bias			RMSE			Inc. Prob.		
Domain	Str.	Sep.	Shr.	Str.	Sep.	Shr.	Str.	Sep.	Shr.
1 (1's)	-0.065	-0.118	-0.063	0.383	0.526	0.385	0.896	0.772	0.899
0's	0.001	0.001	-0.001	0.209	0.283	0.213	0.069	0.185	0.067
2 (1's)	-0.059	-0.106	-0.059	0.381	0.524	0.384	0.900	0.778	0.901
0's	0.001	-0.002	-0.002	0.201	0.280	0.206	0.065	0.181	0.064
3 (1's)	-0.064	-0.116	-0.058	0.380	0.521	0.382	0.898	0.776	0.901
0's	0.001	0.005	0.004	0.201	0.274	0.210	0.066	0.183	0.065
4 (1's)	-0.066	-0.125	-0.064	0.383	0.527	0.381	0.896	0.769	0.900
0's	0.003	0.002	0.000	0.212	0.278	0.214	0.069	0.185	0.066
5 (1's)	-0.060	-0.115	-0.063	0.383	0.524	0.378	0.899	0.773	0.901
0's	-0.001	-0.001	0.002	0.198	0.273	0.216	0.064	0.181	0.067

Table 6.2: Average bias, root mean squared error (RMSE), and inclusion probabilities (Inc. Prob.) for predictors within each Domain, split by those nonzero predictors (1's) and zero predictors (0's), comparing the Structured (Str.), Separate (Sep.), and Shared (Shr.) approaches with homogenous domains

predictive and non-predictive of $\boldsymbol{\mu}^{\mathbf{diff}}$ and $\boldsymbol{\mu}$ for Simulations 1 and 2, respectively.

In Simulation 1 where $\widetilde{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$ and there is domain structure, the Structured method yields lower bias and RMSE, and inclusion probabilities closer to 1 and 0, than the Separate and Shared methods. The Structured method performs particularly well in homogeneous domains (i.e., Domains 1 and 5) where domain-specific effects $\widetilde{\nu}_q$ and ν_q in the inclusion indicators' prior more greatly influenced the inclusion probabilities. The Separate method performs the worst across the board, a sensible result because it treats estimation of $\widetilde{\boldsymbol{\Delta}}$ and $\boldsymbol{\Delta}$ independently when here they are perfectly correlated. A perhaps surprising result is the Structured method outperforming the Shared method in Domain 3, in which half of the variables are predictive of $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\mathbf{diff}}$ and half are not. The Shared method, which imposes the exact same covariate set be included for $\boldsymbol{\mu}$ and $\boldsymbol{\mu}^{\mathbf{diff}}$ with a prior inclusion probability of 1/2, is intuitively the ideal approach here. To investigate this result, we re-ran Simulation 1 with domains separately, and found that the Structured and Shared methods perform equivalently. This suggests that the Structured method, which does a better job of variable selection in Domains 1, 2, 4, and 5, performs better than the Shared method in Domain 3 because the correct variables are included in other domains, reducing the residual noise present in the estimation of Domain 3 parameters.

In Simulation 2 where $\widetilde{\boldsymbol{\Delta}} = \boldsymbol{\Delta}$ but there is no domain structure, the Shared method again seems to be the ideal approach. However, the Structured and Shared methods perform very similarly in terms of bias, RMSE, and inclusion probabilities, suggesting that the Structured method is robust to domain structure. The Separate method performs the worst here for the same reason it performs the worst in Simulation 1. Together, these simulations show that even in scenarios ideally suited for the Shared method, there is no drawback to applying the Structured method.

6.3 Application to 2008-2017 PPCRS

We now apply the model in (6.1) using the priors in (6.2) to the 2008 and 2017 PPCRS, MNCM, and ACS data. Here, the domains $q = 1, ..., Q = 5$ correspond to five domains of primary care structure identified in Bodenheimer and Wagner's Chronic Care Model:[16] clinical information systems (CIS), decision support (DS), delivery system redesign (DSR), health care organization (HCO), and self-management support (SMS). We also adjust the model for differences in clinic-averaged clinical and demographic covariates age, biological sex (=1 if Female, =0 otherwise), indicator of ischemic vascular disease (IVD) (=1 if IVD, =0 otherwise), indicator of Type 1 diabetes (=1 if Type 1, =0 otherwise), indicator of Commercial insurance type (=1 if Commercial, =0 otherwise), indicator of tobacco status (=1 if any tobacco use, =0 otherwise), wealth, and income from 2008 to 2017. We collect these differences in clinic-averaged covariates in a matrix $\boldsymbol{X}$. To get each covariate on a common scale, we transformed the covariates and each outcome to the standard normal scale. The

full model statement using the Structured approach for this application is

$$
\begin{aligned}
&Y_{ji}^{(0)} \sim N(\mu_j, \widetilde{\sigma}_j^2) \\
&Y_{ji}^{(1)} \sim N(\mu_j + \mu_j^{\text{diff}}, \sigma_j^2) \\
&\boldsymbol{\mu} \sim N(\boldsymbol{X}\widetilde{\boldsymbol{\beta}} + \boldsymbol{P}^{\textbf{diff}}\widetilde{\boldsymbol{\Delta}}, \widetilde{\tau}^2\boldsymbol{I_J}) \\
&\boldsymbol{\mu}^{\textbf{diff}} \sim N(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{P}^{\textbf{diff}}\boldsymbol{\Delta}, \tau^2\boldsymbol{I_J}) \\
&\widetilde{\beta}_k \sim (1-\widetilde{w}_k)N(0, 0.025^2) + \widetilde{w}_k t_5(0,5) \\
&\beta_k \sim (1-w_k)N(0, 0.025^2) + w_k t_5(0,5) \\
&\widetilde{\Delta}_{qz} \sim (1-\widetilde{v}_{qz})N(0, 0.0005^2) + \widetilde{v}_{qz}N(0,1^2) \\
&\Delta_{qz} \sim (1-v_{qz})N(0, 0.0005^2) + v_{qz}N(0,1^2) \\
&\widetilde{v}_{qz}^* = \widetilde{\nu}_q + \widetilde{\epsilon}_{qz}, \\
&v_{qz}^* = \nu_q + \epsilon_{qz}, \\
&\begin{bmatrix} \widetilde{\epsilon}_{qz} \\ \\ \epsilon_{qz} \end{bmatrix} \sim \boldsymbol{N_2}(\boldsymbol{0_2}, \boldsymbol{\Omega}), \text{ with } \boldsymbol{\Omega} = \begin{bmatrix} 1 & \rho \\ \\ \rho & 1 \end{bmatrix} \\
&\pi(\widetilde{\nu}_q) = \pi(\nu_q) = N(0,1) \\
&\pi(\rho) = (1-\rho)^{\eta-1}
\end{aligned}
\tag{6.3}
$$

where $k = 1, ..., K$ indexes demographic covariates and $\boldsymbol{P}^{\textbf{diff}} \equiv \boldsymbol{P}^{\textbf{(2017)}} - \boldsymbol{P}^{\textbf{(2008)}}$.

Results

Table 6.3 contains the 95% credible interval estimates for each element in $\widetilde{\boldsymbol{\Delta}}$ and $\boldsymbol{\Delta}$, as well as the inclusion probabilities for each item, for the A1c, LDL, and SBP outcomes. Overall, not a lot of PPCRS items reached statistical significance. Notably, most of the 95% credible intervals have similar bounds, indicating that the vast majority of the items favor the spike component over the slab. Item F3 when using the SBP outcome was the only item with an inclusion probability estimated over 10%. A post-hoc run without any variable selection (i.e., all items were always included) also

		A1c		LDL		SBP	
Item	Domain	$\widehat{v}$	$\boldsymbol{\Delta}$	$\widehat{v}$	$\boldsymbol{\Delta}$	$\widehat{v}$	$\boldsymbol{\Delta}$
B01	CIS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
B02	CIS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)
B06	CIS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
C09	CIS	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.001	(-0.002, 0.002)
C10	CIS	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.000	(-0.002, 0.002)
C14	CIS	0.001	(-0.002, 0.002)	0.002	(-0.002, 0.002)	0.002	(-0.002, 0.002)
C15	CIS	0.002	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.003	(-0.002, 0.002)
C16	CIS	0.001	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.002	(-0.002, 0.002)
B03	DS	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.005	(-0.002, 0.002)
B04	DS	0.001	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
B05	DS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)
C01	DS	0.001	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
C05	DS	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.001	(-0.002, 0.002)
C19	DS	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.000	(-0.002, 0.002)
C31	DS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)
C32	DS	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.003	(-0.002, 0.002)
C33	DS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
C34	DS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
C11	DSR	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.050	(-0.046, 0.002)
C12	DSR	0.000	(-0.002, 0.002)	0.002	(-0.002, 0.002)	0.001	(-0.002, 0.002)
C13	DSR	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)
C17	DSR	0.001	(-0.002, 0.002)	0.017	(-0.003, 0.002)	0.002	(-0.002, 0.002)
C18	DSR	0.000	(-0.002, 0.002)	0.002	(-0.002, 0.002)	0.001	(-0.002, 0.002)
C27	DSR	0.001	(-0.002, 0.002)	0.005	(-0.002, 0.002)	0.001	(-0.002, 0.002)
C28	DSR	0.001	(-0.002, 0.002)	0.022	(-0.003, 0.002)	0.003	(-0.002, 0.002)
C30	DSR	0.001	(-0.002, 0.002)	0.063	(-0.073, 0.002)	0.001	(-0.002, 0.002)
F01	HCO	0.001	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.036	(-0.0309, 0.002)
F02	HCO	0.001	(-0.002, 0.002)	0.002	(-0.002, 0.002)	0.004	(-0.002, 0.002)
F03	HCO	0.001	(-0.002, 0.002)	0.002	(-0.002, 0.002)	**0.178**	(-0.058, 0.002)
C23	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
C24	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
C25	SMS	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.000	(-0.002, 0.002)
C26	SMS	0.000	(-0.002, 0.002)	0.002	(-0.002, 0.002)	0.000	(-0.002, 0.002)
C29	SMS	0.000	(-0.002, 0.002)	0.002	(-0.002, 0.002)	0.007	(-0.002, 0.002)
C35	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D01	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D02	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D03	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D04	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D05	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D06	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D07	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D08	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)
D09	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D10	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D12	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D15	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D16	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D17	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)
D18	SMS	0.000	(-0.002, 0.002)	0.001	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D19	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D20	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D21	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)
D22	SMS	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)	0.000	(-0.002, 0.002)

Table 6.3: Inclusion probabilities and 95% credible intervals of $\boldsymbol{\Delta}$, split by PPCRS question domain

	CIS	DS	DSR	HCO	SMS
A1c	(-2.84, -0.39)	(-2.92, -0.56)	(-2.83, -0.39)	(-2.51, 0.37)	(-3.25, -1.20)
LDL	(-2.84, -0.39)	(-2.92, -0.56)	(-2.80, -0.31)	(-2.51, 0.37)	(-3.25, -1.19)
SBP	(-2.84, -0.38)	(-2.92, -0.55)	(-2.82, -0.34)	(-2.43, 0.63)	(-3.26, -1.19)

Table 6.4: 95% credible intervals of PPCRS question domain-specific effects ν_q for each outcome

yielded highly null results; only one item was significantly predictive of change in A1c and only one item was significantly predictive of change in SBP. Table 6.4 provides the 95% credible intervals for each of the PPCRS question domain-specific effects ν_q for each analyzed outcome. Overall, the the ν_q terms are estimated to be negative, which is sensible because PPCRS items are very rarely selected to be included in the $\boldsymbol{\mu}^{\mathbf{diff}}$ model ($\nu_q = 0$ corresponds to 50% items within domain q included). Within each domain, the credible intervals are very similar across outcomes. Because the outcomes are put on the same scale, this suggests that each outcome had a similarly non-predictive number of items across domains, which is supported by Table 6.3.

Chapter 7

Conclusions

We began this work by attempting to measure the effectiveness of primary care redesign policy on patients living with diabetes in Minnesota. Several characteristics of the data prevented us from applying a standard modeling technique, including the inability to match patients across timepoints, its natural hierarchical structure, and the potential for confounding. With these considerations in mind, we went on to introduce a hierarchical extension of the difference-in-differences model, the causal estimand and assumptions for identification therein, and applied the model to a study on diabetes outcomes and primary care redesign at clinics in Minnesota from 2008 to 2012. While we did not find a significant causal impact, we show that an unadjusted correlational analysis would have overstated evidence of an association between primary care redesign and A1c levels.

The role of covariate adjustment in the HDiD model is not immediately apparent, so we went on to present theoretical and simulation results characterizing the role that covariates play in this context. There, we showed that estimation of the causal effect is biased unless both the baseline and change models are appropriately specified. We then proposed four Bayesian approaches to perform variable selection, assessing their performance via simulation and applying them to measure the impact of primary care redesign on diabetes outcomes in Minnesota from 2008 to 2017. We found that the same principal component that was not found to be significant in the 2008-2011 application *was* significant in the 2008-2017 application, suggesting that primary care redesign lowered cholesterol

and blood pressure among diabetes patients in Minnesota. MN Community Measurement data from 2019 will be available shortly, so a re-application of the HDiD model to measure trends in diabetes outcomes from 2008 to 2019 may yield more significant results than the 2008-2012 analysis of Chapter 3 or the 2008-2017 application of Chapter 4.

While using one principal component reduced the dimensionality of the PPCRS survey, we lost meaningful information about which particular systems of care management most greatly impact diabetes outcome. To recover this information, we applied a Bayesian cross-sectional multi-outcome model to the 2017 PPCRS, MNCM, and ACS data with clinics grouped by their rurality. The distribution of question domains were approximately proportional across rurality subtypes, so we could not discern which mechanisms drive inter-rurality disparities in the relationship between primary care delivery and diabetes outcomes. However, we discovered that the four components from the "D5" optimal diabetes care indicator were nearly independent, suggesting that collapsing them into one single measure may not be optimal.

We concluded this work by proposing a final variable selection algorithm which leverages the temporal structure of the HDiD context, and any natural structure in the predictors. This new "Structured" approach provides a more flexible correlation structure between the inclusion indicators relative to the Separate and Shared approaches of Chapter 4. We show via simulation that when $\widetilde{\boldsymbol{\Delta}}$ and $\boldsymbol{\Delta}$ are perfectly correlated, the Structured method always performs at least as well as the Shared and Separate methods. We applied the Structured approach to the 2008 and 2017 PPCRS and MNCM data, and found very few PPCRS items significantly predictive of A1c, LDL, or SBP. The robustness of the Structured approach may also be well-suited for other settings with structured and high-dimensional predictor spaces, such as statistical genetics.

References

[1] Ryan A, Burgess J Jr and Dimick J. Why we should not be indifferent to specification choices for difference-in-differences. Health Services Research 2015; 50(4).

[2] Card D and Krueger AB. Minimum wages and employment: A case study of the fast-food industry in New Jersey and Pennsylvania. The American Economic Review 1994; 84(4): 772–793.

[3] Angrist J and Pischke J. *Mostly Harmless Econometrics*. Princeton University Press, 2008.

[4] Abraham S and Sung L. Estimating dynamic treatment effects in event studies with heterogenous treatment effects. SSRN Electronic Journal 2018; .

[5] Pietrangelo A and Cherney K. The Effects of Diabetes on Your Body. https://www.healthline.com/health/diabetes/effects-on-body, 2017. [Online; accessed 29-10-2018].

[6] Mayo Clinic. Type 2 diabetes, 2018. http://www.mayoclinic.org/diseases-conditions/type-2-diabetes/ .

[7] American Diabetes Association. Standards of medical care in diabetes. The Journal of Clinical and Applied Research and Education 2018; 41(1).

[8] Berwick D, Nolan T and Whittington J. The triple aim: Care, health, and cost. Health Affairs 2008; 27(3): 759–769.

[9] Blackwell D, Lucas J and Clarke T. Summary health statistics for U.S. adults: National Health Interview Survey, 2012. Vital and Health Statistics Series 10, Data from the National Health Survey 2012; : 1–161.

[10] Shojania K, Ranji S, Shaw L et al. In *Closing the Quality Gap: A Critical Analysis of Quality Improvement Strategies Volume 2: Diabetes Care*. AHRQ Rockville MD, 2004.

[11] Richardson W, Berwick D, Brisgard JC et al. In *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, D.C.: National Academy Press, 2001.

[12] Minnesota Statutes. 2008 Minnesota Session Laws, Chapter 4764, Health Care Homes 2010, 2008.

[13] US Department of Health and Human Services Agency for Healthcare Research and Quality. Defining the PCMH, 2018. URL https://pcmh.ahrq.gov/page/defining-pcmh.

[14] Swaney A. *Quality Measurement Enhancement Project (QMEP)*. Master's thesis, University of Minnesota - Twin Cities, 2018.

[15] Solberg LI, Crain AL, Tillema J et al. Medical home transformation: A gradual process and a continuum of attainment. The Annals of Family Medicine 2013; 11(Suppl 1): S108–S114.

[16] Bodenheimer T, Wagner EH and Grumbach K. Improving primary care for patients with chronic illness. Journal of the American Medical Association 2002; 288(15): 1909 – 1914.

[17] Normington J, Lock E, Carlin C et al. A Bayesian difference-in-difference framework for the impact of primary care redesign on diabetes outcomes. Statistics and Public Policy 2019; 6(1): 55–66.

[18] Rubin D. Estimating causal effects of treatments in randomized and nonrandomized studies. Journal of Educational Psychology 1974; 66(5): 688–701.

[19] Lechner M. The estimation of causal effects by difference-in-difference methods. Foundations and Trends in Econometrics 2010; 4(3): 167–219.

[20] Athey S and Imbens G. Design-based analysis in difference-in-differences settings with staggered adoption. NBER Working Papers 24963 2018; National Bureau of Economic Research, Inc.

[21] de Chaisemartin C and D'Haultfœuille X. Two-way fixed effects estimators with heterogeneous treatment effects. 2018; Unpublished.

[22] Rosenbaum PR and Rubin DB. The central role of the propensity score in observational studies for causal effects. Biometrika 1983; 70(1): 41–55.

[23] Jolliffe IT. Principal component analysis and factor analysis. In *Principal Component Analysis*. Springer, 1986. pp. 115–128.

[24] Lock E, Hoadley K, Marron J et al. Joint and individual variance explained (JIVE) for integrated analysis of multiple data types. Annals of Applied Statistics 2013; 7(1): 523–542.

[25] Hamming R. Error detecting and error correcting codes. The Bell System Technical Journal 1950; 29(2).

[26] Rosenbaum PR. The consequences of adjustment for a concomitant variable that has been affected by the treatment. Journal of the Royal Statistical Society Series A (General) 1984; : 656–666.

[27] Montgomery J, Nyhan B and Torres M. How conditioning on post-treatment variables can ruin your experiment and what to do about it. http://www.dartmouth.edu/~nyhan/post-treatment-bias.pdf, 2017. [Online; accessed 2018-02-14].

[28] Robins J, Hernan MA and Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000; 11(5): 550–560.

[29] Jeffreys H. An invariant form for the prior probability in estimation problems. Proceedings of the Royal Society of London, Series A, Mathematical and Physical Sciences 1946; 186(1007): 453–461.

[30] Spiegelhalter D, Best N, Carlin B et al. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B 2002; 64(4): 583–639.

[31] Geman S and Geman D. Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machine Intelligence 1984; PAMI-6(6): 721–741.

[32] Centers for Disease Control and Prevention. Long-term Trends in Diabetes. https://www.cdc.gov/diabetes/statistics/slides/long_term_trends.pdf, 2017. [Online; accessed 29-10-2018].

[33] Athey S and Imbens G. Identification and inference in nonlinear difference-in-differences models. Econometrica 2006; 74(2): 431–497.

[34] Rosenbaum P and Rubin D. The central role of the propensity score in observational studies for causal effects. Biometrika 1983; 70(1): 41–55. DOI:10.1093/biomet/70.1.41.

[35] Horvitz D and Thompson D. A generalization of sampling without replacement from a finite universe. Journal of the American Statistical Association 1952; 47(260): 663–685. DOI:10.1080/01621459.1952.10483446.

[36] Wang C, Parmigiani G and Dominici F. Bayesian effect estimation accounting for adjustment uncertainty. Biometrics 2012; 68: 661–671.

[37] Cefalu M. Model-averaged double robust estimation. Biometrics 2016; 73(2): 410–421.

[38] Zigler C and Dominici F. Uncertainty in propensity score estimation: Bayesian methods for variable selection and model averaged causal effects. Journal of the American Statistical Association 2014; 109(505): 95–107.

[39] Koch B, Vock D and Wolfson J. Covariate selection with group lasso and doubly robust estimation of causal effects. Biometrics 2017; 74: 8–17.

[40] Wing C, Simon K and Bello-Gomez R. Designing difference in difference studies: Best practices for public health policy research. Annual Review of Public Health 2018; 39(1): 453–469. DOI: 10.1146/annurev-publhealth-040617-013507. PMID: 29328877.

[41] Abadie A. Semiparametric difference-in-differences estimators. Review of Economic Studies 2005; 72: 1–19.

[42] Ionescu-Ittu R and Maria Glymour JK. A difference-in-differences approach to estimate the effect of income-supplementation on food insecurity. Preventive Medicine 2015; 70: 108–116.

[43] McKinnon B, Harper S, Kaufman J et al. Removing user fees for facility-based delivery services: a difference-in-differences evaluation from ten sub-Saharan African countries. Health Policy and Planning 2015; 30(4): 432–441.

[44] Stuart E, Huskamp H, Duckworth K et al. Using propensity scores in difference-in-differences models to estimate the effects of a policy change. Health Services and Outcomes Research Methodology 2014; 14(4): 166–182.

[45] Sofer T, Richardson D, Colicino E et al. On negative outcome control of unobserved confounding as a generalization of difference-in-differences. Statistical Science 2016; 31(3): 348–361.

[46] Pearl J. Causal diagrams for empirical research. Biometrika 1995; 82(4): 669–688.

[47] Textor J, van der Zander B, Gilthorpe MK et al. Robust causal inference using directed acyclic graphs: the R package "DAGitty". International Journal of Epidemiology 2016; 45(6): 1887–1894.

[48] Greene WH. *Econometric Analysis, Fifth Edition*. Prentice Hall, 2003.

[49] Mitchell T and Beauchamp J. Bayesian variable selection in linear regression. Journal of the American Statistical Association 1988; 83: 1023–1032.

[50] George E and McCulloch R. Variable selection via Gibbs sampling. Journal of the American Statistical Association 1993; 88(423): 881–889. DOI:10.1080/01621459.1993.10476353.

[51] Zigler C, Watts K, Yeh R et al. Model feedback in Bayesian propensity score estimation. Biometrics 2013; 69(1): 263–273.

[52] Doob J. Application of the theory of martingales. Coll Int du CNRS, Paris 1948; : 22–28.

[53] Rosenbaum P. The consequences of adjustment for a concomitant variable that has been affected by the treatment. Journal of the Royal Statistical Society Series A 1984; : 656–666.

[54] RUCA map classifications. http://depts.washington.edu/uwruca/ruca-maps.php. [Online; accessed 2019-09-17].

[55] Peterson K, Radosevich D, O'Connor P et al. Improving diabetes care in practice: findings from the TRANSLATE trial. Diabetes Care 2008; 31(12): 2238–2243.

[56] Chew B, Shariff-Ghazili S and Fernandez A. Psychological aspects of diabetes care: Effecting behavioral change in patients. World Journal of Diabetes 2014; 5(6): 796–808. DOI:10.4239/wjd.v5.i6.796.

[57] Albert J and Chib S. Bayesian analysis of binary and polychotomous response data. Journal of the American Statistical Association 1993; 88(422): 669–679.

[58] Cowles MK, Carlin B and Connett J. Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. Journal of the American Statistical Association 1996; 91(433): 86–98.

[59] Murray T, Hobbs B and Carlin B. Combining nonexchangeable functional or survival data sources in oncology using generalized mixture commensurate priors. The Annals of Applied Statistics 2015; 9(3): 1549–1570.

[60] Lewandowski D, Kurowicka D and Joe H. Generating random correlation matrices based on vines and extended onion method. Journal of Multivariate Analysis 2009; 100(9): 1989–2001.

Appendix A

Notation in UNITED Application

Symbol	Meaning
μ	Mean diabetes outcome, 2009
μ^{diff}	Mean change in diabetes outcome from 2009 to 2012
σ^2	Variance of diabetes outcome in both 2009 and 2012
Δ	Effect of clinic structure on mean change in diabetes outcome
$\boldsymbol{\beta}$	Coefficients corresponding to pre- to post-intervention mean adjustment
$\widetilde{\boldsymbol{\beta}}$	Coefficients corresponding to pre-intervention mean adjustment
ψ	Health care system fixed effect for change in mean diabetes outcome
υ	Health care system fixed effect for mean diabetes outcome, 2009
$\widetilde{\tau}^2$	Variance of error term for regression on μ
τ^2	Variance of error term for regression on μ^{diff}
σ^2_{age}	Variance of patient age
θ	True proportion of patients on Commercial insurance
η	True mean age of patients
γ	True proportion of female patients
κ	True proportion of patients with ischemic vascular disease (IVD)
ζ^2	Variance of system effect ψ
ω^2	Variance of system effect υ
$\pi()$	Prior distribution

Table A.1: Notation Reference, Greek Letters

Symbol	Meaning
$^{(yr)}$	Value is specific to year yr
$\sim$	Value is specific to year 2009
$_{sji}$	Value is specific to patient i at clinic j within system s
c	"Clinic score" measuring maturity in primary care transformation
S	Total number of health care systems
J	Total number of clinics
J_s	Total number of clinics within system s
n_{sj}	Total number of patients at clinic j within system s
c^{diff}	Change in clinic score from 2008 to 2011
Y_{sji}	Patient-level diabetes outcome
Comm_{sj}	Sample proportion of patients using commercial insurance
Age_{sj}	Sample mean patient age
Fem_{sj}	Sample proportion of patients who are female
IVD_{sj}	Sample proportion of patients with ischemic vascular disease (IVD)
Wlth_{sj}	Mean of clinic j's patients' neighborhood-level measure of wealth
Inc_{sj}	Mean of clinic j's patients' neighborhood-level measure of income & education
NHW_{sj}	Mean of clinic j's patients' neighborhood-level proportion of non-Hispanic white residents

Table A.2: Notation Reference, Roman Letters

Appendix B

Gibbs Samplers

B.1 2008-2012 UNITED Analysis

The following algorithm outlines the Gibbs sampler used to estimate the model parameters in the 2008-2012 UNITED application of Chapter 3. Here, let $\widetilde{\mathbf{\Theta}} = [\widetilde{\beta}_0,\ \widetilde{\Delta},\ \widetilde{\beta}_1,\ \widetilde{\beta}_2,\ \widetilde{\beta}_3,\ \widetilde{\beta}_4,\ \widetilde{\beta}_5]^T$.

Initialize.

For iterations t=2, ..., T,

For system organization $s = 1, ..., S$,

For clinics $j = 1, ..., J$,

Draw $\sigma_{sj}^{2,(t)} \sim$ IG$(0.5\text{n}_{sj} + \hat{a},$
$0.5(\sum_{i=1}^{\text{n}_{sj}^{(2009)}} (Y_{sji} - \mu_{sj}^{(2009),(t-1)})^2 + \sum_{i=1}^{\text{n}_{sj}^{(2012)}} (Y_{sji}^{(2012)} - \mu_{sj}^{(2009),(t-1)} - \mu_{sj}^{\text{diff},(t-1)})^2) +$
$\hat{b})$, where $\hat{a}$ and $\hat{b}$ are computed via Empirical Bayes by matching the moments of the inverse-gamma distribution.

Draw $\mu_{sj}^{(2009),(t)} \sim$
$N\Bigg(\left[\widetilde{\tau}^{2,(t-1)}(\text{n}_{sj}^{(2009)}\bar{Y}_{sj.}^{(2009)} - \text{n}_{sj}^{(2012)}\mu_{sj}^{\text{diff},(t-1)}) + \sigma_{sj}^{2,(t)}(X_{sj}\widetilde{\mathbf{\Theta}}^{(t-1)} + \upsilon_s^{(t-1)})\right] / \left[\text{n}_{sj}\widetilde{\tau}^{2,(t-1)} + \sigma_{sj}^{2,(t)}\right],$
$\left[(\widetilde{\tau}^{2,(t-1)}\sigma_{sj}^{2,(t)})/(\text{n}_{sj}\widetilde{\tau}^{2,(t-1)} + \sigma_{sj}^{2,(t)})\right]\Bigg).$

Draw $\mu_{sj}^{\text{diff,(t)}} \sim$
$N\Bigg(\left[\text{n}_{sj}^{(2012)}\tau^{2,(t-1)}(\bar{Y}_{sj.}^{(2012)} - \mu_{sj}^{(2009),(t)}) + \sigma_{sj}^{2,(t)}(X_{sj}\mathbf{\Theta}^{(t-1)} + \psi_s^{(t-1)})\right] / \left[\text{n}_{sj}^{((2012))}\tau^{2,(t-1)} + \sigma_{sj}^{2,(t)}\right],$

$\left[\tau^{2,(t-1)}\sigma^{2,(t)}_{sj}\right] / \left[\text{n}^{(2012)}_{sj}\tau^{2,(t-1)} + \sigma^{2,(t)}_{sj}\right]\Bigg)$.

For yr = 2009 and 2012,

Draw $\theta^{(\text{yr}),(t)}_{sj} \sim \text{Beta}(\text{X.Comm}^{(\text{yr})}_{sj} + 1, \text{n}^{(yr)}_{sj} - \text{X.Comm}^{(\text{yr})}_{sj} + 1)$.

Draw $\eta^{(\text{yr}),(t)}_{sj} \sim N\left(\text{Age}^{(\text{yr})}_{sj}, \sigma^{2,(t-1)}_{ji,\text{age}}/\text{n}^{(\text{yr})}_{sj}\right)$.

Draw $\gamma^{(\text{yr}),(t)}_{sj} \sim \text{Beta}(\text{X.Fem}^{(\text{yr})}_{sj} + 1, \text{n}^{(yr)}_{sj} - \text{X.Fem}^{(\text{yr})}_{sj} + 1)$.

Draw $\kappa^{(yr),(t)}_{sj} \sim \text{Beta}(\text{X.IVD}^{(\text{yr})}_{sj} + 1, \text{n}^{(yr)}_{sj} - \text{X.IVD}^{(\text{yr})}_{sj} + 1)$.

Draw $\sigma^{2,(t)}_{sj,age} \sim IG\left(\hat{a}_{\text{age}} + 1, 0.5\text{n}^{(yr)}_{sj}\sum_{yr\in\{2009,2012\}}(\text{Age}^{(\text{yr})}_{sj} - \eta^{(yr),(t)}_{sj})^2 + \hat{b}_{\text{age}}\right)$

Generate new clinic j within system i covariate data with iteration t's parameters according to Section 3.3.3.

Draw $\upsilon^{(t)}_s \sim N\Bigg([\omega^{2,(t-1)}\sum_{j\in s}(\mu^{(2009),(t)}_{sj} - X_{sj}\widetilde{\boldsymbol{\Theta}}^{(t-1)})]/[J\omega^{2,(t-1)} + \widetilde{\tau}^{2,(t-1)}],$
$[\widetilde{\tau}^{2,(t-1)}\omega^{2,(t-1)}]/[J\omega^{2,(t-1)} + \widetilde{\tau}^{2,(t-1)}]\Bigg)$.

Draw $\psi^{(t)}_s \sim N\Bigg([\zeta^{2,(t-1)}\sum_{j\in s}(\mu^{\text{diff},(t)}_{sj} - X_{sj}\boldsymbol{\Theta}^{(t-1)})]/[J\zeta^{2,(t-1)} + \tau^{2,(t-1)}],$
$[\tau^{2,(t-1)}\zeta^{2,(t-1)}]/[J\zeta^{2,(t-1)} + \tau^{2,(t-1)}]\Bigg)$.

Draw $\omega^{2,(t)} \sim IG(0.5S, 0.5\sum_{s=1}^{S}\upsilon^2_s)$.

Draw $\zeta^{2,(t)} \sim IG(0.5S, 0.5\sum_{s=1}^{S}\psi^2_s)$.

Draw $\widetilde{\tau}^{2,(t)} \sim IG\left(0.5J, 0.5\sum_{s=1}^{S}\sum_{j\in s}(\mu^{(2009),(t)}_{sj} - X_{sj}\widetilde{\boldsymbol{\Theta}}^{(t-1)} - \upsilon^{(t)}_s)^2\right)$

Draw $\tau^{2,(t)} \sim IG\left(0.5J, 0.5\sum_{s=1}^{S}\sum_{j\in s}(\mu^{\text{diff},(t)}_{sj} - X_{sj}\boldsymbol{\Theta}^{(t-1)} - \psi^{(t)}_s)^2\right)$

Update $\boldsymbol{\Omega}$ and $\widetilde{\boldsymbol{\Omega}}$ based on the new draws $\tau^{2,(t)}$ and $\widetilde{\tau}^{2,(t)}$.

Draw $\boldsymbol{\Theta}^{(t)} \sim \boldsymbol{N_7}((\boldsymbol{X^T\Omega^{-1}X})^{-1}\boldsymbol{X^T\Omega^{-1}\mu^{\text{diff}}}, (\boldsymbol{X^T\Omega^{-1}X})^{-1})$.

Draw $\widetilde{\boldsymbol{\Theta}}^{(t)} \sim \boldsymbol{N_7}((\boldsymbol{X^T\widetilde{\Omega}^{-1}X})^{-1}\boldsymbol{X^T\widetilde{\Omega}^{-1}\mu}, (\boldsymbol{X^T\widetilde{\Omega}^{-1}X})^{-1})$.

B.2 Variable Selection Simulation Study

This section outlines the Gibbs sampling algorithms used in the simulation studies outlined in Chapter 4. Throughout, let $\boldsymbol{X}$ be the design matrix with a column of 1's for the intercept, $\boldsymbol{T}$, and all candidate covariates so that $\widetilde{\boldsymbol{\Theta}} = [\widetilde{\beta}_0,\ \widetilde{\Delta},\ \widetilde{\beta}_1, ..., \widetilde{\beta}_K]$ and $\boldsymbol{\Theta} = [\beta_0,\ \Delta,\ \beta_1, ..., \beta_K]$.

Separate Method

Initialize.

For iterations $t = 2, ..., T$,

For clinics $j = 1, ..., J$,

Draw $\mu_j^{(t)} \sim N\left(\frac{\widetilde{\tau}^{2,(t-1)}\{\sigma_j^{2,(t-1)} n_j^{(0)} \bar{\boldsymbol{Y}}_{j.}^{(0)} + \widetilde{\sigma}_j^{2,(t-1)} n_j^{(1)} (\bar{\boldsymbol{Y}}_{j.}^{(1)} - \mu_j^{\text{diff},(t-1)})\} + \widetilde{\sigma}_j^{2,(t-1)} \sigma_j^{2,(t-1)} \boldsymbol{X}_j \widetilde{\boldsymbol{\Theta}}^{(t-1)}}{\widetilde{\tau}^{2,(t-1)}(n_j^{(0)} \sigma_j^{2,(t-1)} + n_j^{(1)} \widetilde{\sigma}^{2,(t-1)}) + \widetilde{\sigma}_j^{2,(t-1)} \sigma_j^{2,(t-1)}}, \frac{\widetilde{\sigma}_j^{2,(t-1)} \sigma_j^{2,(t-1)} \widetilde{\tau}^{2,(t-1)}}{\widetilde{\tau}^{2,(t-1)}(n_j^{(0)} \sigma_j^{2,(t-1)} + n_j^{(1)} \widetilde{\sigma}_j^{2,(t-1)}) + \widetilde{\sigma}_j^{2,(t-1)} \sigma_j^{2,(t-1)}} \right)$

Draw $\mu_j^{\text{diff},(t)} \sim N\left(\frac{n_j^{(1)} \tau^{2,(t-1)} (\bar{\boldsymbol{Y}}_{j.}^{(1)} - \mu_j^{(t)}) + \sigma_j^{2,(t-1)} \boldsymbol{X}_j \boldsymbol{\Theta}^{(t-1)}}{n_j^{(1)} \tau^{2,(t-1)} + \sigma_j^{2,(t-1)}}, \frac{\sigma_j^{2,(t-1)} \tau^{2,(t-1)}}{n_j^{(1)} \tau^{2,(t-1)} + \sigma_j^{2,(t-1)}} \right)$

Draw $\widetilde{\sigma}_j^{2,(t)} \sim IG(0.5 n_j^{(0)}, 0.5 \sum_{i=1}^{n_j^{(0)}} (Y_{ji}^{(0)} - \mu_j^{(t)})^2)$

Draw $\sigma_j^{2,(t)} \sim IG(0.5 n_j^{(1)}, 0.5 \sum_{i=1}^{n_j^{(1)}} (Y_{ji}^{(1)} - \mu_j^{(t)} - \mu_j^{\text{diff},(t)})^2)$

Draw $\widetilde{\tau}^{2,(t)} \sim IG(0.5J, 0.5 ||\boldsymbol{\mu}^{(t)} - \boldsymbol{X}\widetilde{\boldsymbol{\Theta}}^{(t-1)}||_2)$

Draw $\tau^{2,(t)} \sim IG(0.5J, 0.5 ||\boldsymbol{\mu}^{\text{diff},(t)} - \boldsymbol{X}\boldsymbol{\Theta}^{(t-1)}||_2)$

Draw $\widetilde{\boldsymbol{\Theta}}^{(t)} \sim \boldsymbol{N}(\{\widetilde{\tau}^{-2,(t)} \boldsymbol{X}^T \boldsymbol{X} + \widetilde{\boldsymbol{D}}^{-2}\}^{-1} \widetilde{\tau}^{-2,(t)} \boldsymbol{X}^T \boldsymbol{\mu}^{(t)}, \{\widetilde{\tau}^{-2,(t)} \boldsymbol{X}^T \boldsymbol{X} + \widetilde{\boldsymbol{D}}^{-2}\}^{-1})$

Draw $\boldsymbol{\Theta}^{(t)} \sim \boldsymbol{N}(\{\tau^{-2,(t)} \boldsymbol{X}^T \boldsymbol{X} + \boldsymbol{D}^{-2}\}^{-1} \tau^{-2,(t)} \boldsymbol{X}^T \boldsymbol{\mu}^{\text{diff},(t)}, \{\tau^{-2,(t)} \boldsymbol{X}^T \boldsymbol{X} + \boldsymbol{D}^{-2}\}^{-1})$

For candidate variables k = 1, ... K,

Draw $\widetilde{w}_k^{(t)} \sim \text{Bern}\left(\frac{1}{1+\widetilde{BF}_k}\right)$, with $\widetilde{BF}_k = \frac{p(\widetilde{\beta}_k^{(t)} | \widetilde{w}_k^{(t)}=0)}{p(\widetilde{\beta}_k^{(t)} | \widetilde{w}_k^{(t)}=1)}$

Draw $w_k^{(t)} \sim \text{Bern}\left(\frac{1}{1+BF_k}\right)$, with $BF_k = \frac{p(\beta_k^{(t)} | w_k^{(t)}=0)}{p(\beta_k^{(t)} | w_k^{(t)}=1)}$

Update $\widetilde{\boldsymbol{D}}$ and $\boldsymbol{D}$ based on the new draws for $\widetilde{\boldsymbol{w}}$ and $\boldsymbol{w}$.

Draw $\widetilde{\gamma}_k^{2,(t)} \sim IG(2.5 + 0.5\widetilde{w}_k^{(t)}, 2.5 * 5^2 + 0.5\widetilde{w}_k^{(t)} \widetilde{\beta}_k^{2,(t)})$

Draw $\gamma_k^{2,(t)} \sim IG(2.5 + 0.5 w_k^{(t)}, 2.5 * 5^2 + 0.5 w_k^{(t)} \beta_k^{2,(t)})$

Shared Method

Initialize.

For iterations $t = 2, ..., T$,

For clinics $j = 1, ..., J$,

Draw $\mu_j^{(t)} \sim N\left(\frac{\widetilde{\tau}^{2,(t-1)}\{\sigma_j^{2,(t-1)} n_j^{(0)} \bar{\boldsymbol{Y}}_{j.}^{(0)} + \widetilde{\sigma}_j^{2,(t-1)} n_j^{(1)} (\bar{\boldsymbol{Y}}_{j.}^{(1)} - \mu_j^{\text{diff},(t-1)})\} + \widetilde{\sigma}_j^{2,(t-1)} \sigma_j^{2,(t-1)} \boldsymbol{X}_j \widetilde{\boldsymbol{\Theta}}^{(t-1)}}{\widetilde{\tau}^{2,(t-1)}(n_j^{(0)} \sigma_j^{2,(t-1)} + n_j^{(1)} \widetilde{\sigma}^{2,(t-1)}) + \widetilde{\sigma}_j^{2,(t-1)} \sigma_j^{2,(t-1)}}, \frac{\widetilde{\sigma}_j^{2,(t-1)} \sigma_j^{2,(t-1)} \widetilde{\tau}^{2,(t-1)}}{\widetilde{\tau}^{2,(t-1)}(n_j^{(0)} \sigma_j^{2,(t-1)} + n_j^{(1)} \widetilde{\sigma}_j^{2,(t-1)}) + \widetilde{\sigma}_j^{2,(t-1)} \sigma_j^{2,(t-1)}} \right)$

Draw $\mu_j^{\text{diff},(t)} \sim N\left(\frac{n_j^{(1)}\tau^{2,(t-1)}(\bar{\boldsymbol{Y}}_{j.}^{(1)}-\mu_j^{(t)})+\sigma_j^{2,(t-1)}\boldsymbol{X}_j\boldsymbol{\Theta}^{(t-1)}}{n_j^{(1)}\tau^{2,(t-1)}+\sigma_j^{2,(t-1)}}, \frac{\sigma_j^{2,(t-1)}\tau^{2,(t-1)}}{n_j^{(1)}\tau^{2,(t-1)}+\sigma_j^{2,(t-1)}}\right)$

Draw $\widetilde{\sigma}_j^{2,(t)} \sim IG(0.5n_j^{(0)}, 0.5\sum_{i=1}^{n_j^{(0)}}(Y_{ji}^{(0)}-\mu_j^{(t)})^2)$

Draw $\sigma_j^{2,(t)} \sim IG(0.5n_j^{(1)}, 0.5\sum_{i=1}^{n_j^{(1)}}(Y_{ji}^{(1)}-\mu_j^{(t)}-\mu_j^{\text{diff},(t)})^2)$

Draw $\widetilde{\tau}^{2,(t)} \sim IG(0.5J, 0.5||\boldsymbol{\mu}^{(t)}-\boldsymbol{X}\widetilde{\boldsymbol{\Theta}}^{(t-1)}||_2)$

Draw $\tau^{2,(t)} \sim IG(0.5J, 0.5||\boldsymbol{\mu}^{\text{diff},(t)}-\boldsymbol{X}\boldsymbol{\Theta}^{(t-1)}||_2)$

Draw $\widetilde{\boldsymbol{\Theta}}^{(t)} \sim \boldsymbol{N}(\{\widetilde{\tau}^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{D}^{-2}\}^{-1}\widetilde{\tau}^{-2,(t)}\boldsymbol{X}^T\boldsymbol{\mu}^{(t)}, \{\widetilde{\tau}^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{D}^{-2}\}^{-1})$

Draw $\boldsymbol{\Theta}^{(t)} \sim \boldsymbol{N}(\{\tau^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{D}^{-2}\}^{-1}\tau^{-2,(t)}\boldsymbol{X}^T\boldsymbol{\mu}^{\text{diff},(t)}, \{\tau^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{D}^{-2}\}^{-1})$

For candidate variables k=1, ..., K,

Draw $w_k^{(t)} \sim \text{Bern}\left(\frac{1}{1+BF_k}\right)$, with $BF_k = \frac{p([\widetilde{\beta}_k^{(t)},\beta_k^{(t)}]^T|w_k^{(t)}=0)}{p([\widetilde{\beta}_k^{(t)},\beta_k^{(t)}]^T|w_k^{(t)}=1)}$

Update $\boldsymbol{D}$ based on the new draw for $\boldsymbol{w}$.

Draw $\gamma_k^{2,(t)} \sim IG(2.5+0.5w_k^{(t)}, 2.5*5^2+0.5w_k^{(t)}(\widetilde{\beta}_k^{2,(t)}+\beta_k^{2,(t)}))$

Sufficient method

Initialize.

For iterations $t = 2, ..., T$,

For clinics $j = 1, ..., J$,

Draw $\mu_j^{(t)} \sim N\left(\frac{\widetilde{\tau}^{2,(t-1)}\{\sigma_j^{2,(t-1)}n_j^{(0)}\bar{\boldsymbol{Y}}_{j.}^{(0)}+\widetilde{\sigma}_j^{2,(t-1)}n_j^{(1)}(\bar{\boldsymbol{Y}}_{j.}^{(1)}-\mu_j^{\text{diff},(t-1)})\}+\widetilde{\sigma}_j^{2,(t-1)}\sigma_j^{2,(t-1)}\boldsymbol{X}_j\widetilde{\boldsymbol{\Theta}}^{(t-1)}}{\widetilde{\tau}^{2,(t-1)}(n_j^{(0)}\sigma_j^{2,(t-1)}+n_j^{(1)}\widetilde{\sigma}^{2,(t-1)})+\widetilde{\sigma}_j^{2,(t-1)}\sigma_j^{2,(t-1)}}, \frac{\widetilde{\sigma}_j^{2,(t-1)}\sigma_j^{2,(t-1)}\widetilde{\tau}^{2,(t-1)}}{\widetilde{\tau}^{2,(t-1)}(n_j^{(0)}\sigma_j^{2,(t-1)}+n_j^{(1)}\widetilde{\sigma}_j^{2,(t-1)})+\widetilde{\sigma}_j^{2,(t-1)}\sigma_j^{2,(t-1)}}\right)$

Draw $\mu_j^{\text{diff},(t)} \sim N(\frac{n_j^{(1)}\tau^{2,(t-1)}(\bar{\boldsymbol{Y}}_{j.}^{(1)}-\mu_j^{(t)})+\sigma_j^{2,(t-1)}\boldsymbol{X}_j\boldsymbol{\Theta}^{(t-1)}}{n_j^{(1)}\tau^{2,(t-1)}+\sigma_j^{2,(t-1)}}, \frac{\sigma_j^{2,(t-1)}\tau^{2,(t-1)}}{n_j^{(1)}\tau^{2,(t-1)}+\sigma_j^{2,(t-1)}})$

Draw $\widetilde{\sigma}_j^{2,(t)} \sim IG(0.5n_j^{(0)}, 0.5\sum_{i=1}^{n_j^{(0)}}(Y_{ji}^{(0)}-\mu_j^{(t)})^2)$

Draw $\sigma_j^{2,(t)} \sim IG(0.5n_j^{(1)}, 0.5\sum_{i=1}^{n_j^{(1)}}(Y_{ji}^{(1)}-\mu_j^{(t)}-\mu_j^{\text{diff},(t)})^2)$

Draw $\widetilde{\tau}^{2,(t)} \sim IG(0.5J, 0.5||\boldsymbol{\mu}^{(t)}-\boldsymbol{X}\widetilde{\boldsymbol{\Theta}}^{(t-1)}||_2)$

Draw $\tau^{2,(t)} \sim IG(0.5J, 0.5||\boldsymbol{\mu}^{\text{diff},(t)}-\boldsymbol{X}\boldsymbol{\Theta}^{(t-1)}||_2)$

Draw $\widetilde{\boldsymbol{\Theta}}^{(t)} \sim \boldsymbol{N}(\{\widetilde{\tau}^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X}+\widetilde{\boldsymbol{D}}^{-2}\}^{-1}\widetilde{\tau}^{-2,(t)}\boldsymbol{X}^T\boldsymbol{\mu}^{(t)}, \{\widetilde{\tau}^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X}+\widetilde{\boldsymbol{D}}^{-2}\}^{-1})$

Draw $\boldsymbol{\Theta}^{(t)} \sim \boldsymbol{N}(\{\tau^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{D}^{-2}\}^{-1}\tau^{-2,(t)}\boldsymbol{X}^T\boldsymbol{\mu}^{\text{diff},(t)}, \{\tau^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X}+\boldsymbol{D}^{-2}\}^{-1})$

For candidate variables k = 1, ... K,

Draw $w_k^e \sim \text{Bern}\left(\frac{1}{1+BF_k^e}\right)$, where $BF_k^e = \frac{N(\alpha_k|0,0.1^2)}{N(\alpha_k|0,1/\gamma_k^e)}$

Draw $w_k \sim w_k^e * \text{Bern}\left(\frac{1}{1+BF_k}\right)$, where $BF_k = \frac{N(\beta_k|0,0.1^2)}{N(\beta_k|0,1/\gamma_k)}$

Draw $\widetilde{w}_k \sim w_k * \text{Bern}\left(\frac{1}{1+\widetilde{BF}_k}\right)$, where $\widetilde{BF}_k = \frac{N(\widetilde{\beta}_k|0,0.1^2)}{N(\widetilde{\beta}_k|0,1/\widetilde{\gamma}_k)}$

Update $\widetilde{\boldsymbol{D}}$ and $\boldsymbol{D}$ based on the new draws for $\widetilde{\boldsymbol{w}}$ and $\boldsymbol{w}$.

Draw $\widetilde{\gamma}_k^{2,(t)} \sim IG(2.5 + 0.5\widetilde{w}_k^{(t)}, 2.5 * 5^2 + 0.5\widetilde{w}_k^{(t)}\widetilde{\beta}_k^{2,(t)})$

Draw $\gamma_k^{2,(t)} \sim IG(2.5 + 0.5w_k^{(t)}, 2.5 * 5^2 + 0.5w_k^{(t)}\beta_k^{2,(t)})$

Efficient method

Initialize.

For iterations $t = 2, ..., T$,

For clinics $j = 1, ..., J$,

Draw $\mu_j^{(t)} \sim N\Bigg(\frac{\widetilde{\tau}^{2,(t-1)}\{\sigma_j^{2,(t-1)} n_j^{(0)} \bar{\boldsymbol{Y}}_{j.}^{(0)} + \widetilde{\sigma}_j^{2,(t-1)} n_j^{(1)} (\bar{\boldsymbol{Y}}_{j.}^{(1)} - \mu_j^{\text{diff},(t-1)})\} + \widetilde{\sigma}_j^{2,(t-1)}\sigma_j^{2,(t-1)}\boldsymbol{X}_j\widetilde{\boldsymbol{\Theta}}^{(t-1)}}{\widetilde{\tau}^{2,(t-1)}(n_j^{(0)}\sigma_j^{2,(t-1)} + n_j^{(1)}\widetilde{\sigma}^{2,(t-1)}) + \widetilde{\sigma}_j^{2,(t-1)}\sigma_j^{2,(t-1)}}, \frac{\widetilde{\sigma}_j^{2,(t-1)}\sigma_j^{2,(t-1)}\widetilde{\tau}^{2,(t-1)}}{\widetilde{\tau}^{2,(t-1)}(n_j^{(0)}\sigma_j^{2,(t-1)} + n_j^{(1)}\widetilde{\sigma}_j^{2,(t-1)}) + \widetilde{\sigma}_j^{2,(t-1)}\sigma_j^{2,(t-1)}}\Bigg)$

Draw $\mu_j^{\text{diff},(t)} \sim N(\frac{n_j^{(1)}\tau^{2,(t-1)}(\bar{\boldsymbol{Y}}_{j.}^{(1)} - \mu_j^{(t)}) + \sigma_j^{2,(t-1)}\boldsymbol{X}_j\boldsymbol{\Theta}^{(t-1)}}{n_j^{(1)}\tau^{2,(t-1)} + \sigma_j^{2,(t-1)}}, \frac{\sigma_j^{2,(t-1)}\tau^{2,(t-1)}}{n_j^{(1)}\tau^{2,(t-1)} + \sigma_j^{2,(t-1)}})$

Draw $\widetilde{\sigma}_j^{2,(t)} \sim IG(0.5n_j^{(0)}, 0.5\sum_{i=1}^{n_j^{(0)}}(Y_{ji}^{(0)} - \mu_j^{(t)})^2)$

Draw $\sigma_j^{2,(t)} \sim IG(0.5n_j^{(1)}, 0.5\sum_{i=1}^{n_j^{(1)}}(Y_{ji}^{(1)} - \mu_j^{(t)} - \mu_j^{\text{diff},(t)})^2)$

Draw $\widetilde{\tau}^{2,(t)} \sim IG(0.5J, 0.5||\boldsymbol{\mu}^{(t)} - \boldsymbol{X}\widetilde{\boldsymbol{\Theta}}^{(t-1)}||_2)$

Draw $\tau^{2,(t)} \sim IG(0.5J, 0.5||\boldsymbol{\mu}^{\text{diff},(t)} - \boldsymbol{X}\boldsymbol{\Theta}^{(t-1)}||_2)$

Draw $\widetilde{\boldsymbol{\Theta}}^{(t)} \sim \boldsymbol{N}(\{\widetilde{\tau}^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X} + \widetilde{\boldsymbol{D}}^{-2}\}^{-1}\widetilde{\tau}^{-2,(t)}\boldsymbol{X}^T\boldsymbol{\mu}^{(t)}, \{\widetilde{\tau}^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X} + \widetilde{\boldsymbol{D}}^{-2}\}^{-1})$

Draw $\boldsymbol{\Theta}^{(t)} \sim \boldsymbol{N}(\{\tau^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{D}^{-2}\}^{-1}\tau^{-2,(t)}\boldsymbol{X}^T\boldsymbol{\mu}^{\text{diff},(t)}, \{\tau^{-2,(t)}\boldsymbol{X}^T\boldsymbol{X} + \boldsymbol{D}^{-2}\}^{-1})$

For candidate variables k = 1, ... K,

Draw $w_k \sim \text{Bern}\left(\frac{1}{1+BF_k}\right)$, where $BF_k = \frac{N(\beta_k|0,0.1^2)}{N(\beta_k|0,1/\gamma_k)}$

Draw $\widetilde{w}_k \sim w_k * \text{Bern}\left(\frac{1}{1+\widetilde{BF}_k}\right)$, where $\widetilde{BF}_k = \frac{N(\widetilde{\beta}_k|0,0.1^2)}{N(\widetilde{\beta}_k|0,1/\widetilde{\gamma}_k)}$

Update $\widetilde{\boldsymbol{D}}$ and $\boldsymbol{D}$ based on the new draws for $\widetilde{\boldsymbol{w}}$ and $\boldsymbol{w}$.

Draw $\widetilde{\gamma}_k^{2,(t)} \sim IG(2.5 + 0.5\widetilde{w}_k^{(t)}, 2.5 * 5^2 + 0.5\widetilde{w}_k^{(t)}\widetilde{\beta}_k^{2,(t)})$

Draw $\gamma_k^{2,(t)} \sim IG(2.5 + 0.5w_k^{(t)}, 2.5 * 5^2 + 0.5w_k^{(t)}\beta_k^{2,(t)})$

B.3 2017 Cross-Sectional Analysis

Let $\boldsymbol{Y} = [\boldsymbol{Y^{(1)}}\ \boldsymbol{Y^{(2)}}\ \boldsymbol{Y^{(3)}}\ \boldsymbol{Y^{*}}]$ where $\boldsymbol{Y^{*}}$ is the latent variable in the probit model for the non-tobacco binary outcome and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, ..., \sigma_D^2)$.

Initialize.

For iterations t = 2, ... T,

For outcomes d = 1, ..., 4,

For subtypes g = 1, 2, 3

Draw $\boldsymbol{\beta_g^{(d)}} \sim \boldsymbol{N}(\boldsymbol{V_d}(\boldsymbol{P_g}^T(\boldsymbol{Y_g^{(d)}} - \boldsymbol{X_g}\boldsymbol{\beta_X^{(d)}} - \boldsymbol{\delta_g^{(d)}}) + \lambda_d^{-2}\bar{\boldsymbol{\beta}}^{(\boldsymbol{d})}), \boldsymbol{V_d})$

where $\boldsymbol{V_d} = (\sigma_d^{-2}\boldsymbol{P_g}^T\boldsymbol{P_g} + \lambda_d{}^{-2}\boldsymbol{I})^{-1}$

Draw $\boldsymbol{\beta_X^{(d)}} \sim \boldsymbol{N}((\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T(\boldsymbol{Y^{(d)}} - \boldsymbol{\theta^{(d)}} - \boldsymbol{\delta^{(d)}}), (\boldsymbol{X}^T\boldsymbol{X})^{-1})$, where $\boldsymbol{\theta_g^{(d)}} = \boldsymbol{P_g}\boldsymbol{\beta}_g^{(\boldsymbol{d})}$

Draw $\bar{\boldsymbol{\beta}}^{(\boldsymbol{d})} \sim N\left(\frac{\boldsymbol{\beta^{(d)}}_0 + \boldsymbol{\beta^{(d)}}_1 + \boldsymbol{\beta^{(d)}}_2}{3}, \frac{\lambda_d^2}{3}\boldsymbol{I}\right)$.

For clinics j = 1, ..., J,

Draw $\delta_{sj}^{(d)} \sim N\left(\frac{\sigma_\delta^{2,(d)}\sum_{i\in j}(Y_{sji}^{(d)} - \boldsymbol{X}_{sji}\boldsymbol{\beta^{(d)}})}{n_{sj}\sigma_\delta^{2,(d)} + \sigma_d^2}, \frac{\sigma_\delta^{2,(d)}\sigma_d^2}{n_{sj}\sigma_\delta^{2,(d)} + \sigma_d^2}\right)$

For each patient i within each clinic j within each system s,

Draw $Y_{sji}^{*}|Y_{sji}^{(4)} = 1 \sim \text{TruncNorm}(\boldsymbol{X}_{sji}^T\boldsymbol{\beta_4} + \delta_{sj}^{(4)}, 1, 0, \infty)$

Draw $Y_{sji}^{*}|Y_{sji}^{(4)} = 0 \sim \text{TruncNorm}(\boldsymbol{X}_{sji}^T\boldsymbol{\beta_4} + \delta_{sj}^{(4)}, 1, -\infty, 0)$, where

TruncNorm(μ, σ, a, b) is the density of the Gaussian distribution with mean μ and standard deviation σ bounded by a on the left and b on the right.

Draw $\boldsymbol{\Sigma}_{11.2} \sim \boldsymbol{IW}(\boldsymbol{A}_{11.2}, N)$, where

$$\boldsymbol{\Sigma}_{11.2} \equiv \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \sigma_{13} \\ \sigma_{12} & \sigma_2^2 & \sigma_{23} \\ \sigma_{13} & \sigma_{23} & \sigma_3^2 \end{bmatrix} - \begin{bmatrix} \sigma_{14}^2 & \sigma_{14}\sigma_{24} & \sigma_{14}\sigma_{34} \\ \sigma_{14}\sigma_{24} & \sigma_{24}^2 & \sigma_{24}\sigma_{34} \\ \sigma_{14}\sigma_{34} & \sigma_{24}\sigma_{34} & \sigma_{34}^2 \end{bmatrix},$$

$$\boldsymbol{A}_{11.2} = (\boldsymbol{Y}_{sji}^{(123)} - \boldsymbol{\mu}_{sji}^{(123)} - -\boldsymbol{\delta}_{sj}^{(123)})(\boldsymbol{Y}_{sji}^{(123)} - \boldsymbol{\mu}_{sji}^{(123)} - \boldsymbol{\delta}_{sj}^{(123)})^T - \frac{(\boldsymbol{Y}_{sji}^{(123)} - \boldsymbol{\mu}_{sji}^{(123)} - -\boldsymbol{\delta}_{sj}^{(123)})(\boldsymbol{Y}_{sji}^{(4)} - \boldsymbol{\mu}_{sji}^{(4)} - \boldsymbol{\delta}_{sj}^{(4)})^T}{(\boldsymbol{Y}_{sji}^{(4)} - \boldsymbol{\mu}_{sji}^{(4)} - -\boldsymbol{\delta}_{sj}^{(4)})^T(\boldsymbol{Y}_{sji}^{(4)} - \boldsymbol{\mu}_{sji}^{(4)} - \boldsymbol{\delta}_{sj}^{(4)})},$$

where $\boldsymbol{\mu}_{sji} = [X_{sji}\boldsymbol{\beta}_1\ X_{sji}\boldsymbol{\beta}_2\ X_{sji}\boldsymbol{\beta}_3\ X_{sji}\boldsymbol{\beta}_4]$ and superscript $^{(\dots)}$ subsets a vector to the listed

diabetes outcomes.

Draw $[\sigma_{14}\ \sigma_{24}\ \sigma_{34}]^T \sim N_3(\frac{1}{a_4^2}[a_{14}\ a_{24}\ a_{34}]^T, \frac{1}{a_4^2}\boldsymbol{\Sigma}_{11.2})$, where

$a_4^2 = (\boldsymbol{Y}^* - \boldsymbol{\mu}^{(4)} - \boldsymbol{\delta}^{(4)})^T(\boldsymbol{Y}^* - \boldsymbol{\mu}^{(4)} - \boldsymbol{\delta}^{(4)})$ and $a_{d4} = (\boldsymbol{Y}^* - \boldsymbol{\mu}^{(d)} - \boldsymbol{\delta}^{(d)})^T(\boldsymbol{Y}^* - \boldsymbol{\mu}^{(4)} - \boldsymbol{\delta}^{(4)})$.

Set $\{\boldsymbol{\Sigma}_y\}_{123,123} = \boldsymbol{\Sigma}_{11.2} + [\sigma_{14}\ \sigma_{24}\ \sigma_{34}]^T[\sigma_{14}\ \sigma_{24}\ \sigma_{34}]$

Set $\{\boldsymbol{\Sigma}_y\}_{4,123} = [\sigma_{14}\ \sigma_{24}\ \sigma_{34}]$.

Set $\{\boldsymbol{\Sigma}_y\}_{123,4} = [\sigma_{14}\ \sigma_{24}\ \sigma_{34}]^T$.

Set $\{\boldsymbol{\Sigma}_y\}_{4,4} = 1$.

Draw $\sigma_\delta^{2,(d)} \sim IG(J/2, \boldsymbol{\delta}^T\boldsymbol{\delta}/2)$.

Draw $\lambda_d^2 \sim IG(3K/2, \frac{1}{2}\sum_{g=1}^3(\bar{\boldsymbol{\beta}}^{(\boldsymbol{d})} - \boldsymbol{\beta}^{(\boldsymbol{d})}{}_g)^T(\bar{\boldsymbol{\beta}}^{(\boldsymbol{d})} - \boldsymbol{\beta}^{(\boldsymbol{d})}{}_g))$

B.4 Structured Method Simulation

Initialize.

For iterations $t = 2, ..., T$,

For groups $j = 1, ..., J$,

Draw $\mu_j^{(t)} \sim N\left(\frac{n_j^{(0)}\bar{\boldsymbol{Y}}_{j.}^{(0)} + n_j^{(1)}(\bar{\boldsymbol{Y}}_{j.}^{(1)} - \mu_j^{\text{diff},(t-1)}) + \boldsymbol{P}_j^T\widetilde{\boldsymbol{\Delta}}^{(t-1)}}{n_j^{(0)}+n_j^{(1)}+1}, \frac{1}{n_j^{(0)}+n_j^{(1)}+1}\right)$

Draw $\mu_j^{\text{diff},(t)} \sim N\left(\frac{n_j^{(1)}(\bar{\boldsymbol{Y}}_{j.}^{(1)} - \mu_j^{(t-1)}) + \boldsymbol{P}_j^T\boldsymbol{\Delta}^{(t-1)}}{n_j^{(1)}+1}, \frac{1}{n_j^{(1)}+1}\right)$

Draw $\widetilde{\boldsymbol{\Delta}}^{(t)} \sim \boldsymbol{N}(\{\boldsymbol{P}^T\boldsymbol{P} + \widetilde{\boldsymbol{G}}^{-2}\}^{-1}\boldsymbol{P}^T\boldsymbol{\mu}^{(t)}, \{\boldsymbol{P}^T\boldsymbol{P} + \widetilde{\boldsymbol{G}}^{-2}\}^{-1})$,

where $\widetilde{\boldsymbol{G}} = \text{diag}[\widetilde{g}_{11}, ..., \widetilde{g}_{Qm_Q}]$, where $\widetilde{g}_{qz} = 5$ if $\widetilde{v}_{qz} = 1$ and $\widetilde{g}_{qz} = 0.1$ if $\widetilde{v}_{qz} = 0$

Draw $\boldsymbol{\Delta}^{(t)} \sim \boldsymbol{N}(\{\boldsymbol{P}^T\boldsymbol{P} + \boldsymbol{G}^{-2}\}^{-1}\boldsymbol{P}^T\boldsymbol{\mu}^{\textbf{diff},(t)}, \{\boldsymbol{P}^T\boldsymbol{P} + \boldsymbol{G}^{-2}\}^{-1})$,

where $\boldsymbol{G} = \text{diag}[g_{11}, ..., g_{Qm_Q}]$, where $g_{qz} = 5$ if $v_{qz} = 1$ and $g_{qz} = 0.1$ if $v_{qz} = 0$

For domains $q = 1, ..., Q$,

For items $z = 1, ..., m_q$,

Draw $[\widetilde{v}_{qz}^{(t)}, v_{qz}^{(t)}]^T \sim \text{Multinomial}(\boldsymbol{p}_{qz}^{(00)}, \boldsymbol{p}_{qz}^{(01)}, \boldsymbol{p}_{qz}^{(10)}, \boldsymbol{p}_{qz}^{(11)})$, where

$$\boldsymbol{p}_{qz}^{(00)} = P(\widetilde{v}_{qz}^{(t)} = 0, v_{qz}^{(t)} = 0) = \frac{\boldsymbol{N_2}([\widetilde{\Delta}_{qz}, \Delta_{qz}]^T, [0,0]^T, \boldsymbol{C_{00}}) \int_{-\infty}^0 \int_{-\infty}^0 \boldsymbol{N_2}([\widetilde{v}_{qz}, v_{qz}]^T, [\widetilde{\nu}_q, \nu_q]^T, \boldsymbol{C_v})}{D_{qz}},$$

$$\boldsymbol{p}_{qz}^{(01)} = P(\widetilde{v}_{qz}^{(t)} = 0, v_{qz}^{(t)} = 1) = \frac{\boldsymbol{N_2}([\widetilde{\Delta}_{qz}, \Delta_{qz}]^T, [0,0]^T, \boldsymbol{C_{01}}) \int_{-\infty}^0 \int_0^\infty \boldsymbol{N_2}([\widetilde{v}_{qz}, v_{qz}]^T, [\widetilde{\nu}_q, \nu_q]^T, \boldsymbol{C_v})}{D_{qz}},$$

$$\boldsymbol{p}_{qz}^{(10)} = P(\widetilde{v}_{qz}^{(t)} = 1, v_{qz}^{(t)} = 0) = \frac{\boldsymbol{N_2}([\widetilde{\Delta}_{qz}, \Delta_{qz}]^T, [0,0]^T, \boldsymbol{C_{10}}) \int_0^\infty \int_{-\infty}^0 \boldsymbol{N_2}([\widetilde{v}_{qz}, v_{qz}]^T, [\widetilde{\nu}_q, \nu_q]^T, \boldsymbol{C_v})}{D_{qz}},$$

$\boldsymbol{p}_{qz}^{(11)} = P(\widetilde{v}_{qz}^{(t)} = 1, v_{qz}^{(t)} = 1) = \frac{\boldsymbol{N_2}([\widetilde{\Delta}_{qz}, \Delta_{qz}]^T, [0,0]^T, \boldsymbol{C_{11}}) \int_0^\infty \int_0^\infty \boldsymbol{N_2}([\widetilde{v}_{qz}, v_{qz}]^T, [\widetilde{\nu}_q, \nu_q]^T, \boldsymbol{C_v})}{D_{qz}}$,

where $\boldsymbol{C_{00}} = \begin{bmatrix} 0.1^2 & 0 \\ 0 & 0.1^2 \end{bmatrix}$, $\boldsymbol{C_{01}} = \begin{bmatrix} 0.1^2 & 0 \\ 0 & 5^2 \end{bmatrix}$, $\boldsymbol{C_{10}} = \begin{bmatrix} 5^2 & 0 \\ 0 & 0.1^2 \end{bmatrix}$, $\boldsymbol{C_{11}} = \begin{bmatrix} 5^2 & 0 \\ 0 & 5^2 \end{bmatrix}$,

$\boldsymbol{C_v} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$, and D_{qz} is the sum of the numerators of $\boldsymbol{p}_{qz}^{(00)}$, $\boldsymbol{p}_{qz}^{(01)}$, $\boldsymbol{p}_{qz}^{(10)}$, and $\boldsymbol{p}_{qz}^{(11)}$.

Draw $\widetilde{v}_{qz}^{*}|\widetilde{v}_{qz}^{(t)} = 1 \sim \text{TruncNorm}(\widetilde{\nu}_q^{(t-1)} + \rho^{(t-1)}(v_{qz}^{*} - \nu_{qz}^{(t-1)}), 1, 0, \infty)$

Draw $\widetilde{v}_{qz}^{*}|\widetilde{v}_{qz}^{(t)} = 0 \sim \text{TruncNorm}(\widetilde{\nu}_q^{(t-1)} + \rho^{(t-1)}(v_{qz}^{*} - \nu_{qz}^{(t-1)}), 1, -\infty, 0)$

Draw $v_{qz}^{*}|v_{qz}^{(t)} = 1 \sim \text{TruncNorm}(\nu_q^{(t-1)} + \rho^{(t-1)}(\widetilde{v}_{qz}^{*} - \widetilde{\nu}_{qz}^{(t-1)}), 1, 0, \infty)$

Draw $v_{qz}^{*}|v_{qz}^{(t)} = 0 \sim \text{TruncNorm}(\nu_q^{(t-1)} + \rho^{(t-1)}(\widetilde{v}_{qz}^{*} - \widetilde{\nu}_{qz}^{(t-1)}), 1, -\infty, 0)$, where

TruncNorm(μ, σ, a, b) is the density of the Gaussian distribution with mean μ and standard deviation σ bounded by a on the left and b on the right.

Draw $\widetilde{\nu}_{qz}^{(t)} \sim N\left(\frac{\sum_{z \in q} \widetilde{v}_{qz}^{*} - \rho^{(t-1)}(v_{qz}^{*} - \nu_q^{(t-1)})}{m_q + 1 - \rho^{2,(t-1)}}, \frac{1}{m_q + 1 - \rho^{2,(t-1)}}\right)$

Draw $\nu_{qz}^{(t)} \sim N\left(\frac{\sum_{z \in q} v_{qz}^{*} - \rho^{(t-1)}(\widetilde{v}_{qz}^{*} - \widetilde{\nu}_q^{(t-1)})}{m_q + 1 - \rho^{2,(t-1)}}, \frac{1}{m_q + 1 - \rho^{2,(t-1)}}\right)$

Update $\widetilde{\boldsymbol{G}}$ and $\boldsymbol{G}$ based on the new draws for $\widetilde{\boldsymbol{v}}$ and $\boldsymbol{v}$, respectively.

Propose $\rho^{*} \sim N(\rho^{(t-1)}, 0.1^2)$.

If $\rho^{*} < -1$ or $\rho^{*} > 1$, discard ρ^{*} and return to previous step.

Compute $r = \frac{h(\rho^{*})}{h(\rho^{(t-1)})}$, where $h(\rho)$ is $(1-\rho^2)^{\eta}$ multiplied by the density of the zero mean bivariate Gaussian distribution with covariance matrix $\begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$.

If $r \geq 1$, set $\rho^{(t)} = \rho^{*}$.

If $r < 1$, set $\rho^{(t)} = \rho^{*}$ with probability r and $\rho^{(t)} = \rho^{(t-1)}$ with probability $1 - r$.

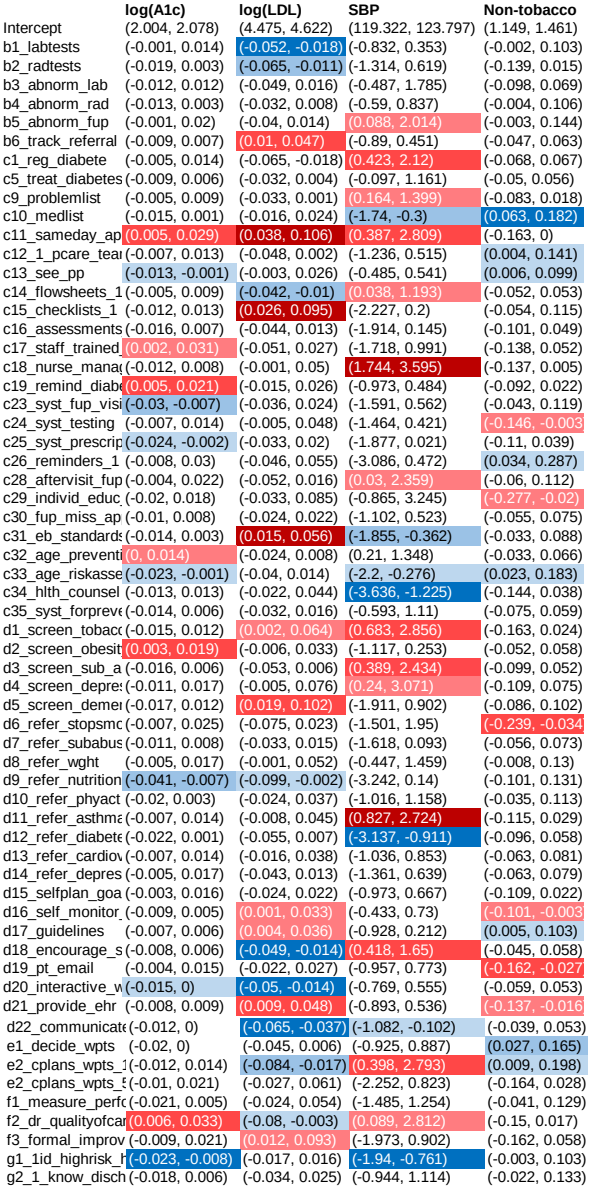
Appendix C

Full Cross-Sectional Analysis Results

Figure C.1: 95% credible intervals from analysis in Chapter 5 for each PPCRS item in *Rural* subgroup; light shades represent significance at 0.05 level, medium shades represent significance at 0.01 level, dark shades represent significance at 0.001 level; blue shades represent desirable diabetes outcomes, red shades represent undesirable diabetes outcomes

	log(A1c)	log(LDL)	SBP	Non-tobacco
Intercept	(1.982, 2.059)	(4.444, 4.568)	(121.704, 125.448)	(1.126, 1.427)
b1_labtests	(-0.003, 0.013)	(-0.012, 0.026)	(-1.869, -0.52)	(0.013, 0.121)
b2_radtests	(-0.017, 0.008)	(-0.057, 0.01)	(1.038, 3.431)	(-0.106, 0.055)
b3_abnorm_lab	(-0.017, 0.008)	(-0.014, 0.056)	(-1.361, 1.11)	(-0.148, 0.016)
b4_abnorm_rad	(-0.001, 0.012)	(-0.018, 0.011)	(-0.1, 0.916)	(-0.038, 0.05)
b5_abnorm_fup	(-0.004, 0.014)	(-0.018, 0.028)	(0.37, 2.003)	(-0.051, 0.076)
b6_track_referral	(-0.015, 0.000)	(-0.025, 0.013)	(-0.947, 0.435)	(-0.054, 0.055)
c1_reg_diabete	(-0.011, 0.009)	(-0.035, 0.017)	(-1.937, -0.085)	(-0.039, 0.097)
c5_treat_diabetes	(-0.006, 0.010)	(-0.002, 0.035)	(-1.872, -0.561)	(-0.01, 0.096)
c9_problemlist	(-0.014, 0.001)	(-0.004, 0.031)	(0.035, 1.265)	(-0.011, 0.084)
c10_medlist	(-0.000, 0.018)	(-0.006, 0.038)	(-0.73, 0.823)	(-0.023, 0.100)
c11_sameday_appt	(0.000, 0.018)	(0.012, 0.057)	(-1.181, 0.422)	(-0.105, 0.02)
c12_1_pcare_teams	(-0.007, 0.01)	(-0.032, 0.012)	(-2.509, -0.963)	(-0.054, 0.069)
c13_see_pp	(-0.011, 0.002)	(-0.005, 0.027)	(1.21, 2.348)	(-0.034, 0.059)
c14_flowsheets_1	(-0.011, 0.003)	(-0.030, 0.001)	(-0.858, 0.217)	(-0.010, 0.082)
c15_checklists_1	(-0.005, 0.021)	(-0.014, 0.054)	(-1.943, 0.505)	(-0.061, 0.108)
c16_assessments_1	(-0.018, 0.006)	(-0.071, 0.001)	(-0.686, 1.772)	(-0.161, -0.001)
c17_staff_trained_1	(-0.003, 0.021)	(-0.044, 0.014)	(-0.801, 1.248)	(-0.093, 0.075)
c18_nurse_manager_1	(-0.01, 0.012)	(0.027, 0.085)	(-0.478, 1.582)	(-0.145, 0.001)
c19_remind_diabete	(0.001, 0.017)	(-0.051, -0.008)	(-0.657, 0.875)	(-0.097, 0.016)
c23_syst_fup_visit	(-0.023, 0.002)	(-0.037, 0.031)	(-2.939, -0.594)	(0.015, 0.182)
c24_syst_testing	(0.001, 0.018)	(-0.005, 0.036)	(-0.23, 1.223)	(-0.102, 0.012)
c25_syst_prescrip	(-0.021, 0.012)	(-0.004, 0.096)	(-0.514, 2.996)	(-0.131, 0.071)
c26_reminders_1	(-0.021, 0.017)	(-0.049, 0.058)	(-3.068, 0.671)	(0.09, 0.348)
c28_aftervisit_fup_1	(-0.012, 0.008)	(-0.056, -0.008)	(-1.231, 0.486)	(-0.065, 0.072)
c29_individ_educ_1	(-0.026, 0.012)	(-0.07, 0.037)	(-1.512, 2.309)	(-0.243, 0.006)
c30_fup_miss_appt_1	(-0.005, 0.01)	(0.004, 0.039)	(-2.337, -1.082)	(-0.189, -0.086)
c31_eb_standards	(-0.015, 0.003)	(-0.022, 0.02)	(-1.849, -0.356)	(-0.079, 0.036)
c32_age_preventives	(-0.001, 0.013)	(-0.001, 0.032)	(0.491, 1.678)	(0.004, 0.102)
c33_age_riskassess	(-0.008, 0.015)	(-0.066, -0.01)	(-0.173, 1.832)	(0.112, 0.268)
c34_hlth_counsel	(-0.013, 0.014)	(-0.054, 0.017)	(-0.140, 2.370)	(-0.143, 0.04)
c35_syst_forprevent	(-0.021, -0.002)	(-0.043, 0.003)	(-1.159, 0.487)	(-0.147, -0.011)
d1_screen_tobacco	(-0.004, 0.023)	(0.016, 0.084)	(0.98, 3.396)	(-0.073, 0.121)
d2_screen_obesity	(-0.002, 0.012)	(-0.008, 0.025)	(-0.681, 0.496)	(-0.091, 0.003)
d3_screen_sub_abuse	(-0.015, 0.009)	(-0.07, -0.001)	(-0.868, 1.471)	(-0.081, 0.072)
d4_screen_depression	(-0.012, 0.011)	(-0.006, 0.053)	(0.348, 2.409)	(-0.105, 0.049)
d5_screen_dementia	(-0.017, 0.009)	(-0.024, 0.044)	(-3.068, -0.637)	(-0.093, 0.079)
d6_refer_stopsmoking	(0.007, 0.032)	(-0.051, 0.01)	(-0.208, 1.952)	(-0.192, -0.03)
d7_refer_subabuse	(-0.004, 0.012)	(0.011, 0.049)	(-0.966, 0.383)	(-0.021, 0.085)
d8_refer_wght	(-0.000, 0.018)	(-0.046, -0.004)	(-1.305, 0.196)	(-0.047, 0.071)
d9_refer_nutrition	(-0.04, -0.01)	(-0.013, 0.07)	(-1.067, 1.88)	(-0.085, 0.127)
d10_refer_phyact	(-0.001, 0.024)	(0.021, 0.093)	(0.517, 2.944)	(-0.085, 0.075)
d11_refer_asthma	(-0.012, 0.011)	(-0.024, 0.035)	(-0.566, 1.52)	(-0.111, 0.038)
d12_refer_diabetes	(-0.02, 0.004)	(-0.027, 0.034)	(-0.847, 1.32)	(-0.087, 0.065)
d13_refer_cardiovas	(-0.002, 0.019)	(0.001, 0.058)	(-0.959, 1.043)	(-0.077, 0.062)
d14_refer_depression	(-0.001, 0.017)	(-0.034, 0.014)	(-1.586, 0.088)	(-0.034, 0.088)
d15_selfplan_goals	(-0.015, 0.001)	(-0.039, 0.001)	(-0.473, 0.975)	(-0.076, 0.038)
d16_self_monitor_help	(-0.008, 0.005)	(0, 0.032)	(-0.083, 1.04)	(-0.072, 0.016)
d17_guidelines	(-0.006, 0.008)	(-0.04, -0.008)	(-1.934, -0.795)	(-0.042, 0.052)
d18_encourage_sm_1	(-0.018, -0.006)	(-0.008, 0.019)	(-1.002, -0.043)	(-0.015, 0.07)
d19_pt_email	(-0.003, 0.014)	(0.006, 0.048)	(-1.395, 0.125)	(-0.128, -0.012)
d20_interactive_web	(-0.016, -0.004)	(-0.048, -0.019)	(-1.147, -0.096)	(0.007, 0.097)
d21_provide_ehr	(-0.010, 0.006)	(-0.031, 0.007)	(-2.307, -0.93)	(-0.151, -0.043)
d22_communicate_ehr	(-0.008, 0.005)	(-0.031, -0.002)	(-0.307, 0.731)	(0.015, 0.103)
e1_decide_wpts	(-0.014, 0.006)	(-0.040, 0.014)	(-1.419, 0.495)	(-0.022, 0.107)
e2_cplans_wpts_1	(-0.017, 0.021)	(-0.119, -0.004)	(-3.132, 0.963)	(-0.102, 0.142)
e2_cplans_wpts_5	(-0.002, 0.016)	(0.003, 0.047)	(-1.535, 0.056)	(-0.125, 0.001)
f1_measure_perform	(-0.029, -0.002)	(0.004, 0.086)	(-2.897, -0.087)	(-0.094, 0.075)
f2_dr_qualityofcare	(-0.007, 0.018)	(-0.082, -0.004)	(0.127, 2.84)	(-0.092, 0.07)
f3_formal_improve	(-0.015, 0.017)	(-0.088, 0.004)	(-1.094, 2.127)	(-0.135, 0.09)
g1_1id_highrisk_hosp	(-0.015, 0.003)	(-0.027, 0.016)	(0.482, 1.994)	(-0.003, 0.11)
g2_1_know_discharge	(0.001, 0.017)	(-0.001, 0.035)	(-0.994, 0.291)	(0.012, 0.119)

Figure C.2: 95% credible intervals from analysis in Chapter 5 for each PPCRS item in *Urban Small* subgroup; light shades represent significance at 0.05 level, medium shades represent significance at 0.01 level, dark shades represent significance at 0.001 level; blue shades represent desirable diabetes outcomes, red shades represent undesirable diabetes outcomes



	log(A1c)	log(LDL)	SBP	Non-tobacco
Intercept	(2.004, 2.078)	(4.475, 4.622)	(119.322, 123.797)	(1.149, 1.461)
b1_labtests	(-0.001, 0.014)	(-0.052, -0.018)	(-0.832, 0.353)	(-0.002, 0.103)
b2_radtests	(-0.019, 0.003)	(-0.065, -0.011)	(-1.314, 0.619)	(-0.139, 0.015)
b3_abnorm_lab	(-0.012, 0.012)	(-0.049, 0.016)	(-0.487, 1.785)	(-0.098, 0.069)
b4_abnorm_rad	(-0.013, 0.003)	(-0.032, 0.008)	(-0.59, 0.837)	(-0.004, 0.106)
b5_abnorm_fup	(-0.001, 0.02)	(-0.04, 0.014)	(0.088, 2.014)	(-0.003, 0.144)
b6_track_referral	(-0.009, 0.007)	(0.01, 0.047)	(-0.89, 0.451)	(-0.047, 0.063)
c1_reg_diabete	(-0.005, 0.014)	(-0.065, -0.018)	(0.423, 2.12)	(-0.068, 0.067)
c5_treat_diabetes	(-0.009, 0.006)	(-0.032, 0.004)	(-0.097, 1.161)	(-0.05, 0.056)
c9_problemlist	(-0.005, 0.009)	(-0.033, 0.001)	(0.164, 1.399)	(-0.083, 0.018)
c10_medlist	(-0.015, 0.001)	(-0.016, 0.024)	(-1.74, -0.3)	(0.063, 0.182)
c11_sameday_ap	(0.005, 0.029)	(0.038, 0.106)	(0.387, 2.809)	(-0.163, 0)
c12_1_pcare_tea	(-0.007, 0.013)	(-0.048, 0.002)	(-1.236, 0.515)	(0.004, 0.141)
c13_see_pp	(-0.013, -0.001)	(-0.003, 0.026)	(-0.485, 0.541)	(0.006, 0.099)
c14_flowsheets_1	(-0.005, 0.009)	(-0.042, -0.01)	(0.038, 1.193)	(-0.052, 0.053)
c15_checklists_1	(-0.012, 0.013)	(0.026, 0.095)	(-2.227, 0.2)	(-0.054, 0.115)
c16_assessments	(-0.016, 0.007)	(-0.044, 0.013)	(-1.914, 0.145)	(-0.101, 0.049)
c17_staff_trained	(0.002, 0.031)	(-0.051, 0.027)	(-1.718, 0.991)	(-0.138, 0.052)
c18_nurse_mana	(-0.012, 0.008)	(-0.001, 0.05)	(1.744, 3.595)	(-0.137, 0.005)
c19_remind_diab	(0.005, 0.021)	(-0.015, 0.026)	(-0.973, 0.484)	(-0.092, 0.022)
c23_syst_fup_visi	(-0.03, -0.007)	(-0.036, 0.024)	(-1.591, 0.562)	(-0.043, 0.119)
c24_syst_testing	(-0.007, 0.014)	(-0.005, 0.048)	(-1.464, 0.421)	(-0.146, -0.003
c25_syst_prescrip	(-0.024, -0.002)	(-0.033, 0.02)	(-1.877, 0.021)	(-0.11, 0.039)
c26_reminders_1	(-0.008, 0.03)	(-0.046, 0.055)	(-3.086, 0.472)	(0.034, 0.287)
c28_aftervisit_fup	(-0.004, 0.022)	(-0.052, 0.016)	(0.03, 2.359)	(-0.06, 0.112)
c29_individ_educ	(-0.02, 0.018)	(-0.033, 0.085)	(-0.865, 3.245)	(-0.277, -0.02)
c30_fup_miss_ap	(-0.01, 0.008)	(-0.024, 0.022)	(-1.102, 0.523)	(-0.055, 0.075)
c31_eb_standards	(-0.014, 0.003)	(0.015, 0.056)	(-1.855, -0.362)	(-0.033, 0.088)
c32_age_preventi	(0, 0.014)	(-0.024, 0.008)	(0.21, 1.348)	(-0.033, 0.066)
c33_age_riskasse	(-0.023, -0.001)	(-0.04, 0.014)	(-2.2, -0.276)	(0.023, 0.183)
c34_hlth_counsel	(-0.013, 0.013)	(-0.022, 0.044)	(-3.636, -1.225)	(-0.144, 0.038)
c35_syst_forprev	(-0.014, 0.006)	(-0.032, 0.016)	(-0.593, 1.11)	(-0.075, 0.059)
d1_screen_tobac	(-0.015, 0.012)	(0.002, 0.064)	(0.683, 2.856)	(-0.163, 0.024)
d2_screen_obesit	(0.003, 0.019)	(-0.006, 0.033)	(-1.117, 0.253)	(-0.052, 0.058)
d3_screen_sub_a	(-0.016, 0.006)	(-0.053, 0.006)	(0.389, 2.434)	(-0.099, 0.052)
d4_screen_depre	(-0.011, 0.017)	(-0.005, 0.076)	(0.24, 3.071)	(-0.109, 0.075)
d5_screen_deme	(-0.017, 0.012)	(0.019, 0.102)	(-1.911, 0.902)	(-0.086, 0.102)
d6_refer_stopsmc	(-0.007, 0.025)	(-0.075, 0.023)	(-1.501, 1.95)	(-0.239, -0.034
d7_refer_subabus	(-0.011, 0.008)	(-0.033, 0.015)	(-1.618, 0.093)	(-0.056, 0.073)
d8_refer_wght	(-0.005, 0.017)	(-0.001, 0.052)	(-0.447, 1.459)	(-0.008, 0.13)
d9_refer_nutrition	(-0.041, -0.007)	(-0.099, -0.002)	(-3.242, 0.14)	(-0.101, 0.131)
d10_refer_phyact	(-0.02, 0.003)	(-0.024, 0.037)	(-1.016, 1.158)	(-0.035, 0.113)
d11_refer_asthm	(-0.007, 0.014)	(-0.008, 0.045)	(0.827, 2.724)	(-0.115, 0.029)
d12_refer_diabet	(-0.022, 0.001)	(-0.055, 0.007)	(-3.137, -0.911)	(-0.096, 0.058)
d13_refer_cardio	(-0.007, 0.014)	(-0.016, 0.038)	(-1.036, 0.853)	(-0.063, 0.081)
d14_refer_depres	(-0.005, 0.017)	(-0.043, 0.013)	(-1.361, 0.639)	(-0.063, 0.079)
d15_selfplan_goa	(-0.003, 0.016)	(-0.024, 0.022)	(-0.973, 0.667)	(-0.109, 0.022)
d16_self_monitor	(-0.009, 0.005)	(0.001, 0.033)	(-0.433, 0.73)	(-0.101, -0.003
d17_guidelines	(-0.007, 0.006)	(0.004, 0.036)	(-0.928, 0.212)	(0.005, 0.103)
d18_encourage_s	(-0.008, 0.006)	(-0.049, -0.014)	(0.418, 1.65)	(-0.045, 0.058)
d19_pt_email	(-0.004, 0.015)	(-0.022, 0.027)	(-0.957, 0.773)	(-0.162, -0.027
d20_interactive_w	(-0.015, 0)	(-0.05, -0.014)	(-0.769, 0.555)	(-0.059, 0.053)
d21_provide_ehr	(-0.008, 0.009)	(0.009, 0.048)	(-0.893, 0.536)	(-0.137, -0.016
d22_communicate	(-0.012, 0)	(-0.065, -0.037)	(-1.082, -0.102)	(-0.039, 0.053)
e1_decide_wpts	(-0.02, 0)	(-0.045, 0.006)	(-0.925, 0.887)	(0.027, 0.165)
e2_cplans_wpts_1	(-0.012, 0.014)	(-0.084, -0.017)	(0.398, 2.793)	(0.009, 0.198)
e2_cplans_wpts_5	(-0.01, 0.021)	(-0.027, 0.061)	(-2.252, 0.823)	(-0.164, 0.028)
f1_measure_perf	(-0.021, 0.005)	(-0.024, 0.054)	(-1.485, 1.254)	(-0.041, 0.129)
f2_dr_qualityofcar	(0.006, 0.033)	(-0.08, -0.003)	(0.089, 2.812)	(-0.15, 0.017)
f3_formal_improv	(-0.009, 0.021)	(0.012, 0.093)	(-1.973, 0.902)	(-0.162, 0.058)
g1_1id_highrisk_h	(-0.023, -0.008)	(-0.017, 0.016)	(-1.94, -0.761)	(-0.003, 0.103)
g2_1_know_disch	(-0.018, 0.006)	(-0.034, 0.025)	(-0.944, 1.114)	(-0.022, 0.133)

Figure C.3: 95% credible intervals from analysis in Chapter 5 for each PPCRS item in *Urban Large* subgroup; light shades represent significance at 0.05 level, medium shades represent significance at 0.01 level, dark shades represent significance at 0.001 level; blue shades represent desirable diabetes outcomes, red shades represent undesirable diabetes outcomes

	log(A1c)	log(LDL)	SBP	Non-tobacco
Intercept	(2.005, 2.073)	(4.422, 4.52)	(119.368, 122.147)	(1.192, 1.463)
b1_labtests	(-0.01, -0.001)	(0.011, 0.029)	(-1.3, -0.673)	(0.033, 0.099)
b2_radtests	(-0.01, 0.002)	(0.017, 0.042)	(-1.066, -0.196)	(-0.039, 0.052)
b3_abnorm_lab	(-0.018, -0.005)	(-0.028, 0)	(-0.563, 0.462)	(-0.091, 0.014)
b4_abnorm_rad	(-0.002, 0.005)	(-0.002, 0.014)	(-0.487, 0.06)	(-0.053, 0.003)
b5_abnorm_fup	(0.006, 0.018)	(-0.027, 0.001)	(-0.343, 0.648)	(-0.004, 0.091)
b6_track_referral	(-0.007, 0)	(0, 0.016)	(0.159, 0.691)	(-0.006, 0.049)
c1_reg_diabete	(-0.008, 0.001)	(-0.021, -0.003)	(-0.888, -0.264)	(-0.025, 0.041)
c5_treat_diabetes	(-0.004, 0.004)	(0.035, 0.049)	(-0.003, 0.49)	(-0.048, 0.006)
c9_problemlist	(-0.006, 0.003)	(-0.017, -0.002)	(0.226, 0.693)	(-0.029, 0.022)
c10_medlist	(-0.012, -0.001)	(-0.03, -0.011)	(-0.045, 0.582)	(0.034, 0.101)
c11_sameday_ap	(0.008, 0.019)	(0.001, 0.025)	(0.453, 1.302)	(-0.104, -0.017)
c12_1_pcare_tea	(-0.01, -0.001)	(-0.007, 0.011)	(-0.177, 0.435)	(-0.035, 0.03)
c13_see_pp	(-0.01, -0.002)	(-0.015, 0)	(-0.65, -0.128)	(0.01, 0.065)
c14_flowsheets_1	(-0.006, 0.001)	(-0.02, -0.007)	(-0.025, 0.42)	(-0.008, 0.04)
c15_checklists_1	(-0.012, 0.001)	(-0.068, -0.041)	(-1.686, -0.699)	(-0.054, 0.045)
c16_assessments	(0, 0.013)	(0.03, 0.054)	(-1.083, -0.267)	(-0.042, 0.042)
c17_staff_trained	(-0.008, 0.01)	(-0.016, 0.018)	(0.885, 2.106)	(-0.213, -0.087)
c18_nurse_mana	(-0.014, -0.004)	(-0.036, -0.016)	(-0.024, 0.665)	(-0.098, -0.026)
c19_remind_diab	(-0.007, 0)	(-0.019, -0.005)	(-0.305, 0.206)	(0.019, 0.073)
c23_syst_fup_visi	(-0.016, 0.006)	(0.029, 0.078)	(0.158, 1.765)	(-0.06, 0.095)
c24_syst_testing	(-0.001, 0.01)	(-0.015, 0.007)	(-0.463, 0.321)	(-0.093, -0.014)
c25_syst_prescrip	(-0.024, -0.005)	(-0.048, -0.013)	(-0.492, 0.656)	(-0.046, 0.068)
c26_reminders_1	(-0.021, 0.013)	(0.026, 0.108)	(-2.908, -0.235)	(0.048, 0.273)
c28_aftervisit_fup	(0.01, 0.03)	(0.028, 0.068)	(1.989, 3.363)	(-0.038, 0.09)
c29_individ_educ	(-0.015, 0.006)	(-0.003, 0.044)	(-1.314, 0.343)	(-0.214, -0.041)
c30_fup_miss_ap	(0.002, 0.011)	(0.027, 0.046)	(1.096, 1.74)	(-0.034, 0.034)
c31_eb_standards	(-0.003, 0.008)	(-0.006, 0.012)	(-1.558, -0.978)	(-0.038, 0.023)
c32_age_preventi	(-0.006, 0.001)	(0.008, 0.022)	(-0.032, 0.466)	(-0.044, 0.009)
c33_age_riskasse	(-0.038, -0.012)	(-0.007, 0.049)	(-0.944, 1.024)	(0.087, 0.25)
c34_hlth_counsel	(0.005, 0.026)	(-0.074, -0.029)	(-0.06, 1.501)	(-0.154, -0.007)
c35_syst_forprev	(-0.003, 0.006)	(-0.022, -0.002)	(-1.924, -1.222)	(-0.046, 0.027)
d1_screen_tobac	(-0.011, 0.025)	(-0.11, -0.02)	(1.458, 4.268)	(-0.148, 0.065)
d2_screen_obesit	(-0.002, 0.008)	(-0.025, -0.009)	(-0.313, 0.175)	(-0.047, 0.005)
d3_screen_sub_a	(-0.005, 0.004)	(-0.017, 0.001)	(0.344, 0.987)	(-0.023, 0.046)
d4_screen_depres	(0.003, 0.019)	(0.03, 0.062)	(0.288, 1.347)	(-0.094, 0.013)
d5_screen_deme	(-0.007, 0.011)	(-0.076, -0.035)	(-0.495, 0.972)	(0.006, 0.147)
d6_refer_stopsmc	(-0.013, 0.006)	(-0.015, 0.024)	(-2.78, -1.352)	(-0.142, -0.005)
d7_refer_subabus	(-0.008, 0.002)	(-0.026, -0.007)	(-1.304, -0.639)	(0.004, 0.072)
d8_refer_wght	(-0.001, 0.014)	(-0.02, 0.009)	(-0.085, 0.891)	(-0.09, 0.005)
d9_refer_nutrition	(-0.023, 0.008)	(0.035, 0.109)	(0.752, 3.467)	(-0.146, 0.075)
d10_refer_phyact	(-0.007, 0.006)	(-0.011, 0.013)	(-0.868, -0.085)	(-0.012, 0.069)
d11_refer_asthm	(-0.007, 0.005)	(-0.025, 0.001)	(0.205, 1.157)	(-0.038, 0.058)
d12_refer_diabet	(-0.005, 0.009)	(-0.028, -0.002)	(0.955, 1.825)	(-0.038, 0.052)
d13_refer_cardio	(0, 0.012)	(-0.01, 0.012)	(0.295, 1.024)	(-0.059, 0.018)
d14_refer_depres	(-0.007, 0.002)	(-0.002, 0.017)	(0.039, 0.74)	(-0.001, 0.07)
d15_selfplan_goa	(-0.006, 0.003)	(-0.008, 0.009)	(0.253, 0.87)	(-0.073, -0.006)
d16_self_monitor	(-0.001, 0.005)	(0.011, 0.024)	(0.365, 0.848)	(-0.019, 0.031)
d17_guidelines	(-0.005, 0.002)	(-0.031, -0.016)	(-1.976, -1.465)	(-0.001, 0.051)
d18_encourage_s	(-0.004, 0.003)	(0.01, 0.023)	(-0.75, -0.278)	(0.003, 0.054)
d19_pt_email	(-0.01, 0.002)	(0, 0.022)	(-0.501, 0.271)	(-0.12, -0.042)
d20_interactive_w	(-0.01, -0.002)	(-0.022, -0.008)	(0.381, 0.889)	(0.018, 0.071)
d21_provide_ehr	(-0.003, 0.006)	(-0.007, 0.011)	(-1.087, -0.438)	(-0.147, -0.079)
d22_communicate	(-0.007, 0)	(-0.005, 0.009)	(-0.467, 0.051)	(0.039, 0.093)
e1_decide_wpts	(0, 0.01)	(0.003, 0.022)	(-1.433, -0.79)	(0.029, 0.096)
e2_cplans_wpts_1	(-0.001, 0.025)	(-0.058, 0.001)	(-0.444, 1.681)	(-0.096, 0.098)
e2_cplans_wpts_5	(0.005, 0.022)	(-0.02, 0.02)	(-2.881, -1.433)	(-0.15, -0.012)
f1_measure_perf	(-0.009, 0.001)	(-0.017, 0.005)	(-1.869, -1.101)	(0.032, 0.115)
f2_dr_qualityofcar	(-0.007, 0.002)	(-0.007, 0.013)	(0.286, 0.999)	(-0.083, -0.006)
f3_formal_improv	(-0.004, 0.021)	(-0.013, 0.043)	(-1.164, 0.818)	(-0.092, 0.095)
g1_1id_highrisk_h	(-0.012, -0.002)	(0.019, 0.04)	(0.047, 0.837)	(-0.013, 0.067)
g2_1_know_disch	(-0.016, 0.002)	(-0.063, -0.022)	(-2.352, -0.93)	(-0.03, 0.099)