The Effects of Cost-Sharing on Colorectal Cancer Screening and Price Shopping*

January 2018

Abstract
Colorectal cancer is the second-leading cause of cancer-related mortality, but the cost-sharing environment for colonoscopies, the preferred form of screening, is complex. This paper combines a machine learning-based double-selection algorithm to perform principled covariate selection with differential exposure to the Affordable Care Act's requirement that insurers fully cover cancer screening services as an instrumental variable to estimate the effect of cost-sharing on colonoscopy utilization and price-shopping. The results imply that every 10 percentage point reduction in patient cost-sharing increases the use of colonoscopies by 0.5 percentage points and the ACA increased the use of colonoscopies by 2.9%, but there is no intensive-margin response. (JEL codes: I12, I13, C14)

Christopher Whaley
RAND Corporation
1776 Main Street
Santa Monica, CA 90401
cwhaley@rand.org

* cwhaley@rand.org The HCCI claims data used in this study are provided by the Health Care Cost Institute and its data contributors—Aetna, Humana, and UnitedHealthcare. Funding provided by NCI R21CA219229-01, the Laura and John Arnold Foundation, and the Bing Center for Health Economics. Helpful comments were provided by seminar participants at RAND.

1. INTRODUCTION

Colorectal cancer is the second-leading cause of cancer-related mortality and approximately 4.5% of individuals will develop colorectal cancer during their lifetimes (Bibbins-Domingo, Grossman, Curry, Davidson, Epling, et al. 2016; ACS 2017). In 2016, 134,000 individuals were diagnosed with and 49,000 individuals died from colorectal cancer in the United States. However, colorectal cancer is highly preventable if detected at an early stage. Colorectal cancers detected before the cancer has spread to lymph nodes are associated with an average survival rate of over 90 percent while cancers detected post-metastasis are associated with a survival rate of only 13 percent (Bretthauer 2011). In recognition of early detection's importance, the U.S. Preventive Services Task Force (USPSTF) recommends that all adults between ages 50 and 75 receive a colonoscopy every 10 years or a stool-based test every three years (Bibbins-Domingo, Grossman, Curry, Davidson, Ebell, et al. 2016). However, adherence to these recommended screening guidelines is poor, and the out-of-pocket costs of screening are a commonly reported barrier to screening (Chao et al. 2004; Denberg et al. 2005). Approximately 32% of eligible adults have never received colorectal cancer screening and 42% have not received screening within the recommended time intervals (Shapiro et al. 2012). Although other tests are available, colonoscopies are the most common form of colorectal cancer detection (Rex et al. 2009).

This paper examines how patient cost-sharing influences patient decisions regarding colonoscopies among the commercially insured population. Due to the importance of colonoscopies as a first-line screening for colorectal cancer, understanding the role of cost-sharing in colonoscopy decision-making is important. Understanding the effects of cost-sharing in this environment is important because cost-sharing mechanisms are a common policy lever to

increase the use of preventive services. However, the cost-sharing environment is complex for several reasons.

First, colonoscopies are among the most expensive cancer screening services. In the nationwide data used in this study, the average colonoscopy price is $1,395. On average, patients were responsible for 12.9% of the colonoscopy's price, $180. Second, colonoscopies are performed for several purposes. They can be performed as screening tests when the patient is not considered to be at risk for colorectal cancer, or for diagnostic purposes when the patient has an existing risk-factor. There are intervention colonoscopies during which the physician removes cancerous polyps, and non-intervention colonoscopies, during which no polyps are detected. Insurance coverage for non-intervention and screening colonoscopies is typically more generous than coverage for intervention and diagnostic colonoscopies. However, patients and their providers often do not know what type of colonoscopy the procedure will be until the colonoscopy is actually conducted. A patient may schedule a routine screening colonoscopy, but then if the physician identifies and removes a polyps during the course of the procedure it will become an intervention colonoscopy and be billed as such. Finally, the Affordable Care Act (ACA) requires that all non-grandfathered insurance plans fully cover all cancer screening services that are recommended as an A or B grade by the USPSTF, including regular colorectal cancer screening. However, the implementation of the ACA's screening requirement has not been uniform among insurers (Pollitz et al. 2012). Some insurers fully cover all colonoscopies and related procedures, while others narrowly define, and cover, only screening and non-intervention colonoscopies. Similarly, coverage mandates do not require insurers to cover cancer screening at all providers and broader cost-sharing trends (e.g. high-deductibles) have increased cost-sharing for many related services (e.g. anesthesia and laboratory tests).

The effect of changes in cost-sharing on patient decisions is not well-understood. While previous studies have examined the effects of state mandates, no studies have directly estimated the effect of changing cost-sharing on patient decisions for colorectal cancer screening (Hamman and Kapinos 2016). Estimating demand elasticities is especially relevant for procedures like colonoscopies, which are one of the most clinically recommended services. While the out-of-pocket costs to patients for colonoscopies can be large, the mortality benefits of colorectal cancer screening are substantial. Thus, the effect of even large changes in cost-sharing for a procedure that can greatly reduce cancer-related mortality is unclear. At the same time, the effects of cost-sharing on patient choice of provider and costs are also not understood.

In this paper, I consider the effects of cost-sharing along two margins: (1) the extensive-margin decision on the choice of whether or not to receive a colonoscopy, and (2) the intensive-margin decision of the price of the colonoscopy, conditional on receiving one, or price shopping.[1] I measure the extensive-margin by examining the effect of cost-sharing on the utilization of colonoscopies. I measure the intensive-margin choice by the price of the colonoscopy procedure. I find modest extensive-margin effects—a 10 percentage point reduction in patient cost-sharing leads to a 0.5 percentage point increase in colonoscopy utilization, which implies an elasticity of 0.16, but no intensive-margin response.

Examining both margins is important because patient cost-sharing is used to limit the moral hazard caused by insurance coverage, which increases the incentives to receive care and may decrease the incentive to price shop. However, in the case of high-value services, like colonoscopies, cost-sharing can lead patients to avoid cost-saving treatments, but the responses

[1] Another intensive-margin exists on the choice of the number of colonoscopies to receive, conditional on receiving at least one colonoscopy. I do not consider this margin, as for patients above age 50 and without elevated risk, colonoscopies are recommended only every 10 years by the USPSTF.

along the intensive margin are less well understood. If patients only respond along the extensive margin, then cost-sharing for high-value services may actually increase spending by increasing utilization without encouraging price shopping (Chernew, Rosen, and Fendrick 2007; Baicker, Mullainathan, and Schwartzstein 2015). However, if patients respond along both the intensive and extensive margins, then insurance designs beyond blunt cost-sharing may be necessary to encourage the use of high-value care while limiting spending (C. M. Whaley, Guo, and Brown 2017). While this paper is limited to a single service, understanding these responses can help inform policies and insurance designs for similar high-value services.

Estimating these relationships faces two primary empirical challenges. First, there is an inherent endogenous relationship between patient cost-sharing and patient choices along each margin. Patients who are more likely to receive a colonoscopy, or a higher-priced colonoscopy, for reasons that are unobservable, might enroll in a more generous insurance plan. To address this endogeneity concern, I instrument for patient cost-sharing using variation in insurance plan generosity combined with a recent policy shift that influences patient cost-sharing for colonoscopies. While the ACA's cancer screening coverage requirements were designed to reduce patient out-of-pocket costs for colonoscopies, many insurance plans fully covered colorectal cancer screening services prior to the ACA. Of the approximately 28,000 insurance plans used in this study, 59.2% generously covered colonoscopies prior to the ACA.

I use the plans that generously covered colonoscopies prior to the ACA as a control group for the plans that did not fully cover colonoscopies prior to the ACA. Among the plans that covered colonoscopy services prior to the ACA, the share of each colonoscopy's total cost that was paid by the patients decreased by an average of less than 1 percentage point following the implementation of the ACA's requirements. For plans that did not fully cover colonoscopies

prior to the ACA, there was an average decrease in the proportion of colonoscopy costs paid by patients of 6.1 percentage points, from 17.8% to 11.7%. Relative to the control plans, patients enrolled in the plans affected by the ACA experienced an approximately 30% reduction in cost sharing for colonoscopies. I use this plan-level variation in the impact of the ACA's cancer screening coverage requirements as an instrument for patient cost-sharing.

The second empirical challenge is the correct choice of covariates to include in the regression model. In this paper, I leverage 2009-2013 data from the Health Care Cost Institute (HCCI). Approximately 50 million individuals from every Metropolitan Statistical Area (MSA) are included in the HCCI data warehouse. An extensive set of data fields, which measure patient, provider, insurance plan, and claim-level characteristics, are included in the data. The comprehensiveness of the data allows for a detailed analysis of the effects of changes in cost-sharing rules on colonoscopy utilization and prices. The breadth of the data allows for the construction of covariates with many functional forms and higher-dimensional covariates of interactions between covariates. However, which of these potential covariates are important to include in the regression model without introducing noise is ambiguous.

To address this question, I use a machine learning-based double-selection approach that is similar to the double-selection Lasso model proposed by Belloni et al. (2012) to select the covariates that are potential confounders of consumer responses along each margin. This model considers all potential covariates, but separates the "signal" from the "noise" by identifying covariates that are predictive of each outcome. I combine the machine-learning variable selection with a post-selection instrumental variables (IV) approach to estimate the causal effect of colonoscopy cost-sharing along each margin (Alexandre Belloni, Chernozhukov, and Hansen 2014a, 2014b; Urminsky, Hansen, and Chernozhukov 2016; Spindler 2016; A. Belloni et al.

2017). While many machine learning approaches are predictive in nature, this approach leverages advances in machine learning approaches to conduct causal inference. The combination of machine learning based-selection with conventional instrumental variables estimation has the properties of selecting covariates in a data-driven fashion while still preserving the useful properties of standard IV estimation approaches.

Using machine-learning techniques to inform covariate selection improves both the model fit and the transparency of the analysis. Estimating treatment effects in a data-driven fashion allows me to examine different covariate structures while reducing concerns about *ex-post* model selection and researcher-driven bias (Ioannidis 2005; Weber et al. 2015). This problem is especially problematic when using a dataset like the HCCI database, which allows for numerous patient characteristics to be included in regression models. One potential solution to this problem is to include all potential covariates in the model. However, this approach may lead to over-fitting, in which case the model coefficients will describe random noise rather than underlying relationships. The second potential solution is to estimate several models and compare the results of each model. However, estimating several models and then selectively choosing the model to report opens the potential for researcher-driven bias. It also opens the possibility that the results may be due to multiple hypothesis testing and thus not representative of the actual responses (Ioannidis 2005; Athey 2015; A. Belloni et al. 2017).

This paper fits into a larger literature on the effects of patient cost-sharing on the use of preventive services. The RAND Health Insurance Experiment first demonstrated that patients are minimally sensitive to changes in out-of-pocket prices, but higher cost-sharing leads to reductions in preventive care (Manning et al. 1987). Insurers often use cost-sharing to reduce the moral hazard of insurance coverage, but in cases where services are particularly high-value, like

colonoscopies, insurers may benefit from reducing, or eliminating cost-sharing (Baicker, Mullainathan, and Schwartzstein 2015). More recent studies have found that reduced cost-sharing modestly increases the use of breast cancer screening mammograms (Trivedi, Rakowski, and Ayanian 2008; Jena et al. 2016). Similar studies have found that "value-based" cost-sharing for prescription drugs increases the use of high-value therapies (Chernew, Rosen, and Fendrick 2007; Chernew et al. 2010; Choudhry et al. 2011). Other studies have used microsimulation models to estimate that waiving cost-sharing for Medicare beneficiaries would increase the use of colonoscopies by 14.6% and thereby reduce colorectal cancer mortality by 13.3% (Peterse et al. 2017). This paper extends this literature by focusing on colorectal cancer screening and by examining both extensive and intensive margin responses. This paper also contributes to the larger evaluation of the insurance reforms introduced by the ACA.

From a methodological perspective, this paper leverages unique data to apply machine learning methods to instrumental variables approaches. Machine learning techniques are widely applied in statistics, computer science, and industry applications and are particularly adept at examining predictive questions. However, they are less widely used in applied econometrics to examine causal questions. This paper provides a practical application of applying machine learning methods to a traditional econometrics question.

The rest of the paper proceeds as follows. Section 2 describes the data used in this study and Section 3 describes the empirical approach. Section 4 discusses the results and Section 5 briefly discuss two robustness tests, which are fully presented in the supplemental Appendix. Section 6 concludes.

2. DATA

2.1. Medical Claims Data

This study uses medical claims data provided by the Health Care Cost Institute (HCCI). The HCCI aggregates medical claims from United Health, Aetna, and Humana and includes claims for approximately 50 million individuals. The HCCI data contains observations from all 50 states and each U.S. metropolitan region. Over the 2009-2013 period used for this study, the HCCI data contains approximately seven million colonoscopy procedures.

The HCCI data includes detailed information on patient demographics (age, sex, industry of occupation, and geographic location), medical utilization, and spending. I use the claims data to measure Charlson comorbidity scores, as well as indicators for each of the 17 chronic conditions that underlie the Charlson score (Charlson et al. 1987). For each patient, I measure the number of inpatient, outpatient, physician, and emergency department visits, and total spending in each setting. I also measure the number of days spent in an inpatient setting. There are 33 first order variables, and they are used as pre-selected covariates in models that serve as comparators to the models that include covariates chosen through machine learning-based double-selection. The HCCI data also allow me to include a wide number of higher-order controls and interactions between different controls. I measure the natural logarithm and square root of each of the spending and utilization measures. I also interact each of the utilization and spending measures, including the non-linear forms, with patient demographics (age, gender), the Charlson scores, each of the 17 chronic condition indicators, insurance type, and patient industry. These alternative functional forms and interactions increase the number of potential covariates for double-selection from 33 to 1,390.

2.1.1. Patient Inclusion and Colonoscopy Identification

Because the USPSTF's colorectal cancer screening recommendations are for individuals between the ages of 50 and 75, all patients included in this study are between the ages of 50 and 64. The HCCI data do not contain Medicare Fee-for-Service claims, so data on the age-65 and above population is unlikely to be reliable and I have excluded these patients. I also exclude any Medicare Advantage enrollees under age 65 and restrict the sample to the commercially insured population. Only patients enrolled in the same insurance plan for the entire calendar year are included in the analytic sample. I further restrict the analysis to plans that are in the HCCI data across all years. For computational reasons, for the extensive-margin analysis, I require all individuals to be in the HCCI data for the entire 2009-2013 period. I include a random 12.5% sample of patients for the colonoscopy utilization analysis, but all patients that meet the inclusion criterion for the intensive-margin analysis.

Colonoscopy procedures are identified according to Current Procedure Terminology (CPT) and International Classification of Disease (ICD-9) codes. Colonoscopies procedures are identified using CPT codes 44388-44394 and 45378-45385 and ICD-9 codes 45.22, 45.23, 45.25, 45.41, and 45.42. Colonoscopy procedures are further differentiated into diagnostic procedures performed on patients with symptoms or conditions that would prompt physician referral, and screening procedures scheduled under routine circumstances.

2.2. Dependent Variable Definitions

For the extensive-margin analysis, I measure the probability that a patient receives a colonoscopy procedure in a given year. The main dependent variable is a binary indicator equal to one if the patient receives a colonoscopy in a year and zero otherwise (Elkin et al. 2012; Jacob et al. 2012). As a sensitivity test, I also measure the probability of receiving a colonoscopy for screening

purposes. For both dependent variables, I do not measure the number of colonoscopies received by each individual within a given year.

For the intensive-margin analysis, I measure the log-transformed price of the colonoscopy procedure's bundled price. The bundled price includes the price of the colonoscopy, plus any other services (e.g. anesthesiology) that were billed as a part of the colonoscopy. This price, commonly referred to as the "allowed amount," captures the amount paid by either the patient, insurer, or employer. It does not include the charge-master price. I winsorize the price outcomes by dropping the top and bottom percentiles of the price distribution.

2.2.1. Colonoscopy Cost-Sharing Definitions

One limitation of the colonoscopy utilization analysis is that colonoscopy cost-sharing is only known for the patients who receive a colonoscopy. Because this analysis examines the decision to receive a colonoscopy, and thus includes patients who do not receive a colonoscopy, the exact cost-sharing for all patients is not known. As a solution, I use the average colonoscopy cost-sharing within each insurance plan and year combination to measure the mean annual cost-sharing within each insurance plan. This measure of expected cost-sharing allows me to include individuals who do not receive a colonoscopy in this analysis, but it also introduces some measurement error as cost-sharing varies within group. For the intensive-margin price shopping analysis, I measure cost-sharing as the proportion of the colonoscopy's cost that was paid by the patient.

3. EMPIRICAL STRATEGY

To estimate the effects of patient cost-sharing along both the extensive (utilization) and intensive (price) margins, I would ideally estimate a regression of the form

$$y_{it} = \alpha + \eta^{OLS} OOP_{it} + g(X_{it}) + \varepsilon_{it} \quad (1)$$

where y_{it} measures either the probability of receiving a colonoscopy or the price of the colonoscopy conditional on receiving one for person i in year t. $g(X_{it})$ includes controls for patient demographics and risk characteristics. Patient cost-sharing is measured by OOP_{it} and thus the η^{OLS}coefficient measures the extensive or intensive margin semi-elasticity with respect to patient cost-sharing, depending on the dependent variable.

This approach faces two empirical challenges. First, there is a possibility of omitted variables bias if the error term is correlated with OOP_{it}. If patients with unobserved preferences for more procedures or higher-priced procedures select more generous health insurance plans, then this endogeneity will cause the η^{OLS} coefficient to be biased. Second, the appropriate structure of $g(X_{it})$ is unclear. To address these empirical challenges, I combine an instrumental variables estimation approach with a variable selection model that uses a machine-learning algorithm to identify the relevant coefficients in $g(X_{it})$.

3.1. ACA's Coverage Requirement as an Instrumental Variable

As a solution to the first empirical challenge, endogeneity between cost-sharing and patient choices, I use the change in cost-sharing caused by the ACA's mandate as an instrumental variable. In particular, I exploit the variation in pre-ACA coverage of colonoscopy services. Many plans fully covered colonoscopies prior to the ACA. For these plans, the ACA's mandate had no impact on patient costs for colonoscopy services. However, the ACA reduced patient costs for colonoscopies dramatically for plans that did not cover colonoscopies prior to the ACA.

This variation in the effectiveness of the ACA creates a natural experiment to examine the effects of changes in cost-sharing on colonoscopy decisions. I apply this natural experiment as an instrumental variable and estimate the following system of equations:

$$OOP_{it} = \alpha + \delta_{DD}\boldsymbol{ACA_t}\times\boldsymbol{treatment}_{ij} + \tau ACA_t + \gamma treatment_{ij} + g(X_{it}) + \mu_{ijt} \quad (2)$$

$$colonoscopy_{it} = \alpha + \eta\widehat{OOP}_{jt} + \tau ACA_t + \gamma treatment_{ij} + g(X_{it}) + \varepsilon_{ijt} \quad (3)$$

In this expression, the $\boldsymbol{ACA_t}\times\boldsymbol{treatment}_{ij}$ instrument interacts the treatment plans with the 2011 implementation of the ACA. In the second stage equation, the dependent variable $colonoscopy_{it}$ measures the response to cost-sharing using the predicted change in cost-sharing from the first stage for both the utilization and price-shopping outcomes. Thus, under the assumptions discussed below, the η coefficient measures the causal effect of cost-sharing on each colonoscopy outcome.

To define $treatment_{ij}$, and thus construct the instrument, I use each plans' change in patient cost-sharing before and after the ACA. I first measure the median patient cost-sharing in each plan before and after the implementation of the ACA's coverage requirements, which went into effect in 2011. I define the treatment plans as the plans where the median patient cost-sharing was positive prior to 2011, but decreased following 2011. The control plans are those where median patient cost-sharing remained zero in both the pre- and post-ACA periods. Of the 27,713 plans included in this analysis, 11,303 are classified as treatment plans and 16,410 are classified as control plans. As a robustness test, Appendix A uses the continuous pre-ACA cost-sharing in each plan as an alternative instrument. The robustness results are similar to the main results.

To be a valid instrument, two assumptions must be met. First, the instrument must lead to changes in patient cost-sharing, $cov\left(OOP_{ijt}, \boldsymbol{ACA_t}\times\boldsymbol{treatment}_{ij}\right) \neq 0$. The differences in cost-sharing between the two plan types, and the relevance of the instrument, are shown in Figure 1, which plots annual patient cost-sharing trends for each plan type for all colonoscopies and screening colonoscopies, respectively. Prior to the ACA, patients were on average

responsible for 17.8% of colonoscopy costs in plans that did not generously cover colonoscopies, but only 9.2% in plans that did cover colonoscopies. Following the ACA's implementation, there was a 6.1 percentage point, equivalent to a 34.3%, reduction in colonoscopy cost-sharing in the non-covered plans. For the treatment plans, patient cost-sharing decreased from 10.6% to 8.9%. For screening colonoscopies, there was a 42.6% reduction from 16.6% to 7.1% in patient cost-sharing among the treatment plans, and a 15.3% reduction from 10.6% to 8.9% in the control plans. Appendix Figure A3.1 plots similar trends for colonoscopies performed for screening purposes.

The second necessary assumption is that conditional on the covariates, the instrument is not correlated with unobserved factors that lead to changes in colonoscopy decision, $cov(\varepsilon_{ijt}, \boldsymbol{ACA_t} \times \boldsymbol{treatment}_{ij}) = 0$. In other words, interpreting the IV regression results causally requires the assumption that individuals with unobserved preferences for colonoscopies, or low-cost colonoscopies, do not differentially enroll in either the treatment or control plans following the 2011 implementation of the ACA's coverage requirements. The assumption that individuals with unobserved preferences do not ever differentially enroll in different types of plans is not required. Patients undoubtedly have unobserved preference for care, which impacts insurance plan choice. The insurance plan fixed effects capture time-invariant preferences for plans.

Because the implementation of the ACA is quasi-randomly implemented, this assumption is likely satisfied. Moreover, as discussed in Section 5.1, patient demographics, risk and chronic condition characteristics, medical spending, and health care utilization do not change between the treatment and control plans, before and after the ACA. If the ACA has led to changes in unobserved patient preferences between the treatment and control plans, then these other patient

characteristics are likely to change as well. Finally, Appendix 2 tests for parallel trends in patient cost-sharing and each colonoscopy outcome, and does not find meaningful pre-ACA differences between the treatment and control plans.

3.2. Machine Learning Algorithm to Improve Covariate Selection

The second empirical challenge is that the set of covariates that should be included in $g(X_{it})$ is not obvious. Most studies rely on economic intuition or institutional knowledge to formulate the appropriate set of controls. However, this type of researcher-driven selection raises the possibility of misspecification if there are errors in the researcher's intuition or institutional knowledge. Many studies address this uncertainty by estimating sensitivity tests that include differing covariate structures. However, this introduces the possibility of researcher-driven bias as to what results to report (Ioannidis 2005; Athey 2015).

While variables such as comorbidities, utilization, and age are likely correlated with colonoscopy decisions, it is not clear which of these covariates, and their specifications, should be included in the regression model. Furthermore, as mentioned above, the breadth of the HCCI data allows for the construction of control variables with several functional forms and higher-order controls that interact different patient characteristics (e.g. age interacted with the number of inpatient days). These higher-order covariates are potentially important to include because the patient characteristics, and heterogeneity in these characteristics, that influence colonoscopy decisions are not well understood. However, it is not obvious which of the higher-order covariates are important to include.

To include these covariates in the model, one approach is to include each potential covariate in $g(X_{it})$. However, this approach may introduce unwanted noise into the model. With the large sample size included in this study, it will also potentially introduce computational

difficulties. Instead, I use cross-validated machine-learning methods to inform the covariate selection and the composition of $g(X_{it})$.

First, I expand $g(X_{it})$ into each potential covariate, x_{ik}, and an error-term specific to each potential covariate, r_{ik} (Alexandre Belloni, Chernozhukov, and Hansen 2014a, 2014b):

$$g(X_{it}) = \sum \beta_k x_{ik} + r_{ik}. \quad (4)$$

The challenge is to identify the covariates that predictively contribute to the model (i.e. $\beta_k \neq 0$), rather than introduce noise (i.e. $\beta_k = 0$). I apply a variant of the Least Absolute Shrinkage and Selection Operator (LASSO) algorithm to estimate which potential covariates have non-zero β coefficients (Tibshirani 1996). I fit a generalized linear model applying penalized maximum likelihood using the GLMnet algorithm (Friedman, Hastie, and Tibshirani 2010). Like LASSO, the GLMnet algorithm iteratively tests potential covariates and identifies the predictive strength of each potential covariate. To avoid including all potential covariates, and thus over-fitting, each covariate is fitted with the penalty term.

The GLMnet algorithm fits a generalized linear model with an elastic net penalty. The elastic net penalty includes l_1 (LASSO) and l_2 penalties (ridge regression), and their mixtures. Thus, the elastic net is able to combine the advantageous properties of the lasso and ridge regression algorithms. It is also computationally more efficient than either algorithm estimated separately.

For each potential covariate, β, the GLMnet fits the predictive value as

$$\min_{(\beta_0,\beta)} \left[\frac{1}{2N} \sum_{i=0}^{N} (y_i - \beta_0 - x_i^T \beta)^2 + \lambda P_\alpha(\beta) \right]. \quad (5)$$

As with the LASSO and ridge regressions, the $\lambda P_\alpha(\beta)$ term represents the penalty factor for each potential covariate. For a given potential covariate, β, the elastic net penalty is given by

$$P_\alpha(\beta) = (1-\alpha)\frac{1}{2}\|\beta\|_{l_2}^2 + \alpha\|\beta\|_{l_1}^2. \quad (6)$$

Under LASSO, $\alpha = 1$, while in a ridge regression, $\alpha = 0$. The GLMnet algorithm identifies the optimal value of α between, and including, 0 and 1.

I fit the GLMnet model using the glmnet package in R (Friedman, Hastie, and Tibshirani, 2014). To estimate the prediction model, for both analyses, I pull a random sample of 100,000 observations. I use a 20-fold cross-validation approach to estimate the model. The cross-validation helps improve both the computational efficiency and accuracy of the model.

3.3. Post-Selection IV Estimation

To apply the machine-learning variable selection model to the instrumental variables estimation, one approach is to simply apply the variable selection model to the reduced form equation (1). However, this approach will still potentially suffer from omitted variables bias if predictors that are weakly or not predictive of the outcome and correlated with the treatment are not included. Instead, I estimate the GLMnet variable selection model on both equation (1), the reduced form equation, and equation (2), the first stage (Alexandre Belloni, Chernozhukov, and Hansen 2014a, 2014b). I include the union of covariates that are selected as predictors of the outcomes in the reduced form equation and first stage equations. In other words, $g(X_{it})$ for models predicting colonoscopy utilization and price includes covariates that are predictive of the respective colonoscopy outcome and the change in cost-sharing following the implementation of the ACA's coverage requirement. Thus, $g(X_{it})$ includes controls for potential confounders of both the outcomes and the endogenous cost-sharing treatment.

With just the selected covariates included in $g(X_{it})$, I then estimate the IV regressions using two-stage least-squares. I apply some measure of economic intuition and iteratively add insurance plan fixed effects. Models that include the insurance plan fixed effects are the

preferred specification, and the plan fixed effects negate the need for the main treatment plan controls. I cluster all standard errors at the employer plan level.

3.4. Model Comparison

After performing the model selection, I fit parallel sets of models. The first set includes just the 33 first order pre-selected covariates, while the second set includes the covariates chosen through machine learning-based double-selection algorithm. I use the same set of pre-selected covariates for both the utilization and price-shopping outcomes. However, the double-selected covariates are specific to each of the two outcomes. For each model set, I compare the fit statistics, specifically R-squared values, from these two models sets to evaluate the potential advantages of the double-selection approach.

4. RESULTS

4.1. Patient Characteristics

Table 1 presents the descriptive characteristics of the patients included in the colonoscopy utilization analysis. This table shows the characteristics of patients enrolled in the treatment and control plans, before and after the ACA's 2011 effective date. The table presents differences in patient demographics, risk scores, chronic condition diagnoses, medical spending, and health care utilization. For each population and time period, the table shows the mean and standard deviations for both plans, and the standardized mean differences between the two plans. Prior to the 2011 implementation of the ACA's coverage requirements, patients in both plans are descriptively very similar. For none of the descriptive characteristics is there a 0.1 standard deviation difference between the two populations, and for most characteristics, there is less than

0.01 standard deviation difference. The differences between the patient populations along these characteristics is almost unchanged following the implementation of the ACA's requirement.

The primary exception is for colonoscopy cost-sharing. Prior to the ACA, patient cost-sharing in the treatment plans was 0.63 standard deviations higher than in the control plans. When measured as a share of the procedure's cost, the treatment plans had 0.78 standard deviations higher cost-sharing. Following the implementation of the ACA's cancer screening coverage requirement, cost-sharing remained almost unchanged for the control plans, but fell considerably for the treatment plans. However, even following the ACA, patient cost-sharing is 0.41 and 0.52 standard deviations higher in the treatment plans than in the control plans, respectively.

Table 2 presents similar results for the intensive-margin analysis. The patients in the treatment and control plans have similar demographics, risk profiles, medical spending, and utilization profiles, both before and after the ACA's requirement. However, as in the previous table, patient cost-sharing for colonoscopies varies substantially. While total colonoscopy costs are similar for both populations and in both time periods, in the pre-period, patients in the treatment plans pay 0.28 and 0.37 standard deviations more for colonoscopies, when measured in dollars and as a share of the procedure's total price, respectively. In the post-ACA period, these differences decrease to 0.12 and 0.15 standard deviations, respectively.

For both the extensive and intensive-margin analyses, the lack of meaningful changes for the patient characteristics other than those related to colonoscopies, supports the exclusion restriction assumption, which is required for the IV results to be interpreted causally. If the reduction in colonoscopy cost-sharing induced by the ACA leads patients who are unobservably responsive to cost-sharing to switch plans, then there are likely to be differences in patient

demographics, other forms of health care spending, and utilization. The absence of these changes suggests that the ACA's cancer screening coverage requirements did not lead patients to switch plans.

4.2. Variable Selection

Figures 2 – 3 present the cross-validated mean-squared errors (MSE) against the inclusion of additional covariates, and the penalty for the inclusion of additional covariates. Figure 2 presents the cross-validation fit for the extensive-margin colonoscopy utilization analysis, while Figure 3 presents the fit for the intensive-margin colonoscopy cost analysis. In each figure, panel A presents the variable selection results for the endogenous treatment confounder selection and panel B presents the reduced form variable selection results.

In Figure 2, adding additional covariates does not reduce the MSE. After a $\log(\lambda) = -8$, the additional covariates introduce substantial noise into the model. These results imply that naively including the covariates which only introduce bias and do not contribute to the predictive ability of the model, will substantially increase the MSE. The largest MSE is 300, while MSE for the selected covariates is less than one. In Figure 3, adding additional controls improves the MSE fit. Around $\log(\lambda) = -6$, the additional controls increase MSE, and in many cases, lead to a rapid increase in MSE. However, the magnitude of the MSE, and thus the introduction of bias due to extraneous covariates, is not nearly as large as in Figure 2.

The GLMnet algorithm selects the covariates that minimize the MSE for each outcome. Thus, out of the approximately 1,400 potential covariates, the model is able to separate those that contribute to the predictive accuracy of the model with those that only add random noise. From the variable selection algorithm, a total of 219 covariates are included for the colonoscopy utilization analysis and 592 covariates are included for the colonoscopy price-shopping analysis.

In the pre-specified specifications, 33 covariates are included. These covariates were chosen based on previous studies and common characteristics that are likely to influence patient colonoscopy decisions.

4.3. First Stage Results—Effect of Exposure to ACA on Patient Cost-Sharing

Table 3 presents the first-stage results that measure the effect of exposure to the ACA's requirements on patient cost-sharing for colonoscopy services. Columns 1 and 2 present results that include the econometrician-selected covariates, while Columns 3 and 4 present the results that include the algorithm-selected covariates. For each set of controls, columns 2 and 4 iteratively add insurance plan fixed effects.

In the top panel, there is a sizable reduction in plan-level patient cost-sharing for the treatment plans relative to the control plans. Regardless of the covariate choice or the inclusion of plan fixed, effects, there is a 3.9 percentage point reduction in the mean patient cost-sharing within each plan. For screening colonoscopies, the bottom panel shows an identical reduction in cost-sharing. Based on the pre-ACA mean plan-level cost-sharing of 21.3% in the treatment plans, these reductions imply an 18.3% reduction in plan-level patient cost-sharing for colonoscopies.

Panel B of Table 3 presents similar results for procedure-level cost-sharing, which is used for the price-shopping analysis. Following the implementation of the ACA's coverage requirement, cost-sharing decreased by approximately 5.2 percentage points in the treatment plans compared to the control plans. Based on the pre-ACA means for the treatment plans of 17.8%, the first-stage effects translate into a 29.2% reduction in patient cost-sharing, which is equivalent to approximately $58.7.

The two sets of first-stage results are slightly different from each other because the results in Panel A use plan-level patient cost-sharing for colonoscopies, while the results in Panel B use patient-level cost-sharing. Out of all specifications, the smallest excluded instrument F test is 54, which indicates that leveraging the changes in cost-sharing following the ACA is a strong instrument (Stock and Yogo 2005). Consistent with the descriptive trends, these results show that the variation in the impact of the ACA, which is based on variation in pre-ACA plan generosity, creates a strong instrument for patient cost-sharing.

4.4. Reduced Form Results—Effect of Exposure to ACA on Colonoscopy Utilization and Price Shopping

Table 4 presents the reduced form results for colonoscopy utilization and price shopping. These results can be interpreted as the difference-in-differences change in each outcome that is due to the ACA's cancer screening coverage requirements. In Panel A, which examines the extensive margin, the preferred specification in column 4 shows that the reduction of patient costs caused by the ACA's requirements led to an 0.22 percentage-point increase in colonoscopy utilization. Based on the baseline utilization rate of 7.7% among the treatment plan population, this result implies that exposure to the ACA increased the use of colonoscopies by 2.9%. The effect is not meaningfully different when using the pre-selected or algorithm-selected covariates, although the coefficients are estimated more precisely when using the algorithm-selected covariates. The effect is slightly smaller for screening colonoscopies, 0.16 percentage points, and the results are again similar in when using the pre-selected and algorithm-selected covariates.

In all four columns, the main ACA effect is negative. These results imply that there is a general time trend towards fewer colonoscopies for both populations. Because the ACA coefficients are larger in magnitude than the ACA x treatment plan variable, the overall

colonoscopy utilization rate decreased for all patients, but to a smaller extent for the individuals in the treatment plan. In addition, the $\boldsymbol{ACA_t \times treatment_{ij}}$ coefficient is similar in magnitude to the -0.002 coefficient for the main treatment plan effect in columns 1 and 3. Thus, the ACA's coverage requirements increased colonoscopy utilization among the treatment plans to approximately the same level of utilization in the more generous control plans.

Panel B presents the reduced-form results for colonoscopy prices. These results do not show any statistically significant change in colonoscopy prices following the implementation of the ACA's cancer screening coverage requirements. The main treatment plan coefficient is negative, which suggests that the treatment plan enrollees receive less expensive colonoscopies, and the main ACA coefficient is positive, which is natural given rising health care prices over time.

While the coefficients are similar when using the pre-selected and algorithm-selected covariates, the model fit is much higher in the double-selection regressions. For colonoscopy utilization, the R^2 in column 2, the preferred pre-selected model, is 0.046. In column 4, which uses the double-selection algorithm to select covariates, the R^2 is 0.147. Similarly, in the intensive-margin analysis, the R^2 in the pre-selected specification is 0.143, compared to 0.58 in the double-selection specification. Thus, including the appropriate high-dimensional covariates leads to a 3-4 times improvement in the amount of variation in colonoscopy decisions that is explained by the model.

4.5. Association-Level Results

Before examining the instrumented results, I present the non-instrumented results that measure the association between patient cost-sharing and each colonoscopy outcome. As shown in Panel A of Table 5, there is a negative association between plan-level cost sharing and colonoscopy utilization. The association is largest when including the insurance plan fixed effects, and

indicates that every 10 percentage-point increase in cost-sharing is associated with a 0.19 percentage point reduction in colonoscopy utilization. The magnitude of the association is much smaller when including the algorithm-selected covariates than when including the pre-selected covariates.

Panel B presents similar results for the association between patient cost-sharing and colonoscopy prices. The pre-selected coefficients indicate a large association between patient cost-sharing and prices. The decrease in cost-sharing caused by the ACA is associated with a 0.8% increase in colonoscopy costs. However, when using the algorithm-selected covariates, there is only a 0.3% increase in costs.

4.6. IV Results

Table 6 estimates the causal effect of patient cost-sharing on both the extensive- and intensive-margin outcomes. The results in the top panel show a negative relationship between cost-sharing and the use of colonoscopy services. The results imply that every 10-percentage point increase in cost-sharing for colonoscopy services reduces the annual use of colonoscopies by approximately 0.5 percentage points. Based on the 18.2% reduction in plan-level cost-sharing estimated in the first stage, this coefficient implies an extensive-margin elasticity of 0.16. This elasticity is close to the RAND Health Insurance elasticity of 0.2, and suggests that changes to cost-sharing have a modest effect on the utilization of colonoscopy services, despite the clinical importance of colorectal cancer screening (Manning et al. 1987).

Notably, the coefficients in the models that do not include covariates, include the intuition-selected covariates, and include the algorithm-selected covariates are almost identical. The post-selection results are slightly larger in magnitude in the algorithm-selected covariate specifications. The cost-sharing coefficients are also estimated more precisely. Not surprisingly,

the R-squared values are much higher in the double-selection specifications. From a methodological perspective, the similarity of the results implies that, at least in this case, the post-selection procedure to identify the most appropriate covariate structure does not substantially change the interpretation of the results compared to traditional instrumental variables results. The independence of the 2SLS results from the covariate specification also supports the strength of the instrument, as the covariates are likely correlated with unobserved patient characteristics.

In addition, although all of the instrumented results are much smaller than the non-causal association-level results in Table 7, there is more similarity for the double-selection results than for the pre-selected specifications. This difference suggests that the double-selection process is able to identify more, but not all, potential confounders than the pre-selection process. This result is not surprising, as the double-selection algorithm is designed to identify the most relevant confounders in the data.

Panel B examines the intensive-margin relationship between patient costs and procedure costs. When using the pre-selected covariates, the preferred results in column 4 show an inverse relationship between patient cost-sharing and procedure costs, albeit the results are not statistically significant. However, this effect disappears in column 6, which includes the algorithm-selected covariates. When including the high-dimensional covariates, these results imply that changing patient cost-sharing has no effect on the cost of colonoscopy procedures.

At least in this setting and among the commercially insured population, there is little evidence of intensive-margin moral hazard for colorectal cancer screening services. Changing cost-sharing appears to have no effect on patient intensive margin choices. The machine learning-based algorithm selection model does lead to meaningfully different coefficient

magnitudes, although the results are not statistically significant in any of the specifications. As in the extensive-margin analysis, the results for the double-selection specifications are much more similar to the OLS results than the pre-selected IV specifications.

5. ROBUSTNESS TESTS

5.1. Continuous Baseline Cost-Sharing as an Instrument

The primary instrumental variables approach used in this paper dichotomizes insurance plans based on variation in exposure to the ACA's coverage mandate. As an alternative instrument, I apply the continuous variation in exposure to the ACA's requirements. Appendix 1 presents these alternative results, which are similar to the main results. The effect of cost-sharing on colonoscopy utilization is slightly smaller, and there is still no effect of cost-sharing on price shopping.

5.2. Parallel Trends Test

This paper's identification approach uses differential exposure to the ACA's requirement that insurance plans fully cover colorectal cancer screening services as an instrumental variable. The use of difference-in-differences variation in cost-sharing as an instrument assumes that cost-sharing only changes due to the differential exposure to the ACA's requirements. One way to test this assumption is to test for pre-ACA trend differences between the treatment and control plan. Any pre-trend differences suggest that other factors might be driving the observed changes in cost-sharing and colonoscopy outcomes. As a test of this assumption, I estimate the first stage and reduced form regressions, but replace the $\boldsymbol{ACA_t \times treatment_{ij}}$ variable with interactions of the treatment plans and the years 2010, 2011, 2012, and 2013. 2010 is the year prior to the ACA's implementation, and so the $2010_t \times treatment_{ij}$ coefficient gives the pre-ACA trend

difference between the treatment and control plans. As shown in Appendix 2, there is very little evidence of pre-trend differences in either the first stage or reduced form regressions.

5.3. Screening Colonoscopies

A final robustness test examines if the responses for colonoscopies performed solely for screening purposes are different than the main results. The results are nearly identical to the main results. As shown in Appendix 3, there is still no intensive-margin price response and the extensive-margin utilization response elasticity is 0.156.

6. CONCLUSION

Colonoscopies are an important service that substantially improve colorectal cancer outcomes. Many policies recognize the importance of colonoscopies, yet the cost-sharing environment for colonoscopies is complex. This study examines the effects of changes in patient cost-sharing but separates extensive and intensive margin responses to cost-sharing. I find that changes in cost-sharing lead to a small change in the use of colonoscopies. However, I do not find that changing patient cost-sharing leads to a change in price-shopping. These results suggest that, at least for colonoscopies, patients are sensitive to cost-sharing along the extensive margin, but do respond along the intensive margin.

To estimate these effects, this study uses variation in exposure to the ACA's requirement that insurers fully cover cancer screening services as an instrumental variable. This study applies a machine learning-based double-selection approach to optimally select covariates among many high-dimensional covariates. Compared to a pre-selected set of covariates, the double-selection approach improves model fit and precision, but does not change the cost-sharing coefficients of

interest. With a strong instrument, as in the case of this study, the specification of covariates does not substantially change the results.

These two types of patient responses have important implications for policy. The ultimate aim of the ACA's mandate, and other similar polices, is to increase the use of colonoscopies, an extensive margin change. The results of this study support the use of cost-sharing mechanisms to increase the use of colorectal cancer screening, but imply that the response is likely to be relatively modest. In this setting, a 30% reduction in patient costs increased the use of colonoscopies, which are one of the most highly recommended forms of preventive care, by approximately 3%. Targeted interventions that reward patients for receiving screening, or information incentives that promote the health benefits of colonoscopies may better increase the use of colonoscopies and similar services. At the same time, the lack of any intensive margin response implies that increasing coverage generosity for colonoscopies is not likely to lead patients to receive care from more expensive providers. While this paper focuses on a single service, better understanding these responses is important to designing insurance policies to encourage the use of other high-value services.

This study has two main limitations. First, while the approach uses a double-selection approach to identify covariates, the selected covariates still come from the claims data. This approach does not improve upon the ability to control patient characteristics that are not observable in the claims data, such as patient family history of cancer, income, and education, but may be correlated with colonoscopy decisions. Second, this study only examines responses to changes in the generosity of cost-sharing among individuals with commercial insurance. Other important factors, such as gaining access to insurance or access to a primary care provider, may

be more important determinants of colorectal cancer screening. Future work should examine the effects of these other pathways.

Nonetheless, the results of this study show how a large set of patients respond to changes in cost-sharing for an important cancer screening service. In 2010, the year prior to the ACA, 52.4% of the study-inclusion eligible individuals in the HCCI population were in a treatment plan, and were thus impacted by the ACA's cancer screening coverage requirement. Nationwide, approximately 43.7 million individuals aged 50-64 were enrolled in commercial insurance in 2010 ("Medical Expenditure Panel Survey" 2017). Following the USPFTF guidelines, approximately 4.37 million of these individuals should receive a colonoscopy in each year. Assuming that the HCCI data is representative of the broader U.S. population, these results imply that exposure to the ACA's reforms led to 50,400 additional patients, or approximately 1% of the eligible population, receiving a colonoscopy. Based on an average colonoscopy cost of $1,395, this increase in colonoscopy utilization implies a spending increase of $70.3 million. However, due largely to the ability to easily treat early-stage cancers, colonoscopies are cost-saving compared to not screening for colorectal cancer (Zauber et al. 2009). While the effect of these screenings on health will be delayed, this increase in colonoscopy use may reduce future colorectal cancer incidence and mortality.

7. REFERENCES

ACS. 2017. "Colorectal Cancer Facts & Figures 2017-2019." Cancer Stat Facts. Atlanta, GA: American Cancer Society.

Athey, Susan. 2015. "Machine Learning and Causal Inference for Policy Evaluation." In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 5–6. KDD '15. New York, NY, USA: ACM. https://doi.org/10.1145/2783258.2785466.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein. 2015. "Behavioral Hazard in Health Insurance." *The Quarterly Journal of Economics* 130 (4):1623–67. https://doi.org/10.1093/qje/qjv029.

Belloni, A., V. Chernozhukov, I. Fernández-Val, and C. Hansen. 2017. "Program Evaluation and Causal Inference With High-Dimensional Data." *Econometrica* 85 (1):233–98. https://doi.org/10.3982/ECTA12723.

Belloni, Alexandre, Victor Chernozhukov, and Christian Hansen. 2014a. "Inference on Treatment Effects after Selection among High-Dimensional Controls." *The Review of Economic Studies* 81 (2):608–50. https://doi.org/10.1093/restud/rdt044.

———. 2014b. "High-Dimensional Methods and Inference on Structural and Treatment Effects." *Journal of Economic Perspectives* 28 (2):29–50. https://doi.org/10.1257/jep.28.2.29.

Bibbins-Domingo, Kirsten, David C. Grossman, Susan J. Curry, Karina W. Davidson, Mark Ebell, John W. Epling, Francisco A. R. García, et al. 2016. "Screening for Skin Cancer: US Preventive Services Task Force Recommendation Statement." *JAMA* 316 (4):429. https://doi.org/10.1001/jama.2016.8465.

Bibbins-Domingo, Kirsten, David C. Grossman, Susan J. Curry, Karina W. Davidson, John W. Epling, Francisco A. R. García, Matthew W. Gillman, et al. 2016. "Screening for Colorectal Cancer: US Preventive Services Task Force Recommendation Statement." *JAMA* 315 (23):2564–75. https://doi.org/10.1001/jama.2016.5989.

Bretthauer, M. 2011. "Colorectal Cancer Screening." *Journal of Internal Medicine* 270 (2):87–98. https://doi.org/10.1111/j.1365-2796.2011.02399.x.

Chao, Ann, Cari J. Connell, Vilma Cokkinides, Eric J. Jacobs, Eugenia E. Calle, and Michael J. Thun. 2004. "Underuse of Screening Sigmoidoscopy and Colonoscopy in a Large Cohort of US Adults." *American Journal of Public Health* 94 (10):1775–81. https://doi.org/10.2105/AJPH.94.10.1775.

Charlson, Mary E., Peter Pompei, Kathy L. Ales, and C. Ronald MacKenzie. 1987. "A New Method of Classifying Prognostic Comorbidity in Longitudinal Studies: Development and Validation." *Journal of Chronic Diseases* 40 (5):373–83. https://doi.org/10.1016/0021-9681(87)90171-8.

Chernew, Michael E., Iver A. Juster, Mayur Shah, Arnold Wegh, Stephen Rosenberg, Allison B. Rosen, Michael C. Sokol, Kristina Yu-Isenberg, and A. Mark Fendrick. 2010. "Evidence That Value-Based Insurance Can Be Effective." *Health Affairs*, January, 10.1377/hlthaff.2009.0119. https://doi.org/10.1377/hlthaff.2009.0119.

Chernew, Michael E., Allison B. Rosen, and A. Mark Fendrick. 2007. "Value-Based Insurance Design." *Health Affairs* 26 (2):w195–203. https://doi.org/10.1377/hlthaff.26.2.w195.

Choudhry, Niteesh K., Jerry Avorn, Robert J. Glynn, Elliott M. Antman, Sebastian Schneeweiss, Michele Toscano, Lonny Reisman, et al. 2011. "Full Coverage for Preventive

Medications after Myocardial Infarction." *The New England Journal of Medicine* 365 (22):2088–97. https://doi.org/10.1056/NEJMsa1107913.
Denberg, Thomas D., Trisha V. Melhado, John M. Coombes, Brenda L. Beaty, Kenneth Berman, Tim E. Byers, Alfred C. Marcus, John F. Steiner, and Dennis J. Ahnen. 2005. "Predictors of Nonadherence to Screening Colonoscopy." *Journal of General Internal Medicine* 20 (11):989–95. https://doi.org/10.1111/j.1525-1497.2005.00164.x.
Elkin, Elena B., Ephraim Shapiro, Jacqueline G. Snow, Ann G. Zauber, and Marian S. Krauskopf. 2012. "The Economic Impact of a Patient Navigator Program to Increase Screening Colonoscopy." *Cancer* 118 (23):5982–88. https://doi.org/10.1002/cncr.27595.
Friedman, Jerome, Trevor Hastie, and Rob Tibshirani. 2010. "Regularization Paths for Generalized Linear Models via Coordinate Descent." *Journal of Statistical Software* 33 (1):1–22.
Hamman, Mary K., and Kandice A. Kapinos. 2016. "Colorectal Cancer Screening and State Health Insurance Mandates." *Health Economics* 25 (2):178–91. https://doi.org/10.1002/hec.3132.
Ioannidis, John P. A. 2005. "Why Most Published Research Findings Are False." *PLoS Med* 2 (8):e124. https://doi.org/10.1371/journal.pmed.0020124.
Jacob, Binu J., Rahim Moineddin, Rinku Sutradhar, Nancy N. Baxter, and David R. Urbach. 2012. "Effect of Colonoscopy on Colorectal Cancer Incidence and Mortality: An Instrumental Variable Analysis." *Gastrointestinal Endoscopy* 76 (2):355–364.e1. https://doi.org/10.1016/j.gie.2012.03.247.
Jena, Anupam B., Jie Huang, Bruce Fireman, Vicki Fung, Scott Gazelle, Mary Beth Landrum, Michael Chernew, Joseph P. Newhouse, and John Hsu. 2016. "Screening Mammography for Free: Impact of Eliminating Cost Sharing on Cancer Screening Rates." *Health Services Research*, March. https://doi.org/10.1111/1475-6773.12486.
Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz. 1987. "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment." *The American Economic Review* 77 (3):251–77. https://doi.org/10.2307/1804094.
"Medical Expenditure Panel Survey." 2017. Rockville, MD: Agency for Healthcare Research and Quality. www.hcup-us.ahrq.gov/faststats/landing.jsp.
Peterse, Elisabeth F. P., Reinier G. S. Meester, Andrea Gini, Chyke A. Doubeni, Daniel S. Anderson, Franklin G. Berger, Ann G. Zauber, and Iris Lansdorp-Vogelaar. 2017. "Value Of Waiving Coinsurance For Colorectal Cancer Screening In Medicare Beneficiaries." *Health Affairs* 36 (12):2151–59. https://doi.org/10.1377/hlthaff.2017.0228.
Pollitz, Karen, Kevin Lucia, Katie Keith, Robert Smith, Mary Doroshenk, Holly Wolf, and Thomas K. Weber. 2012. "Coverage of Colonoscopies Under the Affordable Care Act's Prevention Benefit." Kaiser Family Foundation. http://kff.org/health-costs/report/coverage-of-colonoscopies-under-the-affordable-care/.
Rex, Douglas K., David A. Johnson, Joseph C. Anderson, Phillip S. Schoenfeld, Carol A. Burke, and John M. Inadomi. 2009. "American College of Gastroenterology Guidelines for Colorectal Cancer Screening 2008." *The American Journal of Gastroenterology* 104 (3):739–50. https://doi.org/10.1038/ajg.2009.104.
Shapiro, Jean A., Carrie N. Klabunde, Trevor D. Thompson, Marion R. Nadel, Laura C. Seeff, and Arica White. 2012. "Patterns of Colorectal Cancer Test Use, Including CT Colonography, in the 2010 National Health Interview Survey." *Cancer Epidemiology,*

Biomarkers & Prevention: A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology 21 (6):895–904. https://doi.org/10.1158/1055-9965.EPI-12-0192.

Spindler, Martin. 2016. "Lasso for Instrumental Variable Selection: A Replication Study." *Journal of Applied Econometrics* 31 (2):450–54. https://doi.org/10.1002/jae.2432.

Stock, James, and Motohiro Yogo. 2005. "Testing for Weak Instruments in Linear IV Regression." In *Identification and Inference for Econometric Models*, edited by Donald W. K. Andrews, 80–108. New York: Cambridge University Press. http://www.economics.harvard.edu/faculty/stock/files/TestingWeakInstr_Stock%2BYogo.pdf.

Tibshirani, Robert. 1996. "Regression Shrinkage and Selection via the Lasso." *Journal of the Royal Statistical Society. Series B (Methodological)* 58 (1):267–88.

Trivedi, Amal N., William Rakowski, and John Z. Ayanian. 2008. "Effect of Cost Sharing on Screening Mammography in Medicare Health Plans." *New England Journal of Medicine* 358 (4):375–83. https://doi.org/10.1056/NEJMsa070929.

Urminsky, Oleg, Christian Hansen, and Victor Chernozhukov. 2016. "Using Double-Lasso Regression for Principled Variable Selection." SSRN Scholarly Paper ID 2733374. Rochester, NY: Social Science Research Network. https://papers.ssrn.com/abstract=2733374.

Weber, Ann M., Van Der Laan, Mark J, and Maya L. Petersen. 2015. "Assumption Trade-Offs When Choosing Identification Strategies for Pre-Post Treatment Effect Estimation: An Illustration of a Community-Based Intervention in Madagascar." *Journal of Causal Inference* 3 (1):109–30.

Whaley, Christopher M., Chaoran Guo, and Timothy T. Brown. 2017. "The Moral Hazard Effects of Consumer Responses to Targeted Cost-Sharing." *Journal of Health Economics* 56 (Supplement C):201–21. https://doi.org/10.1016/j.jhealeco.2017.09.012.

Zauber, Ann G., Amy B. Knudsen, Carolyn M. Rutter, Iris Lansdorp-Vogelaar, James E. Savarino, Marjolein van Ballegooijen, and Karen M. Kuntz. 2009. *Cost-Effectiveness of CT Colonography to Screen for Colorectal Cancer*. AHRQ Technology Assessments. Rockville (MD): Agency for Healthcare Research and Quality (US). http://www.ncbi.nlm.nih.gov/books/NBK284750/.

8. TABLES AND FIGURES

Figure 1: Trends in Colonoscopy Cost-Sharing by Insurance Plan Type

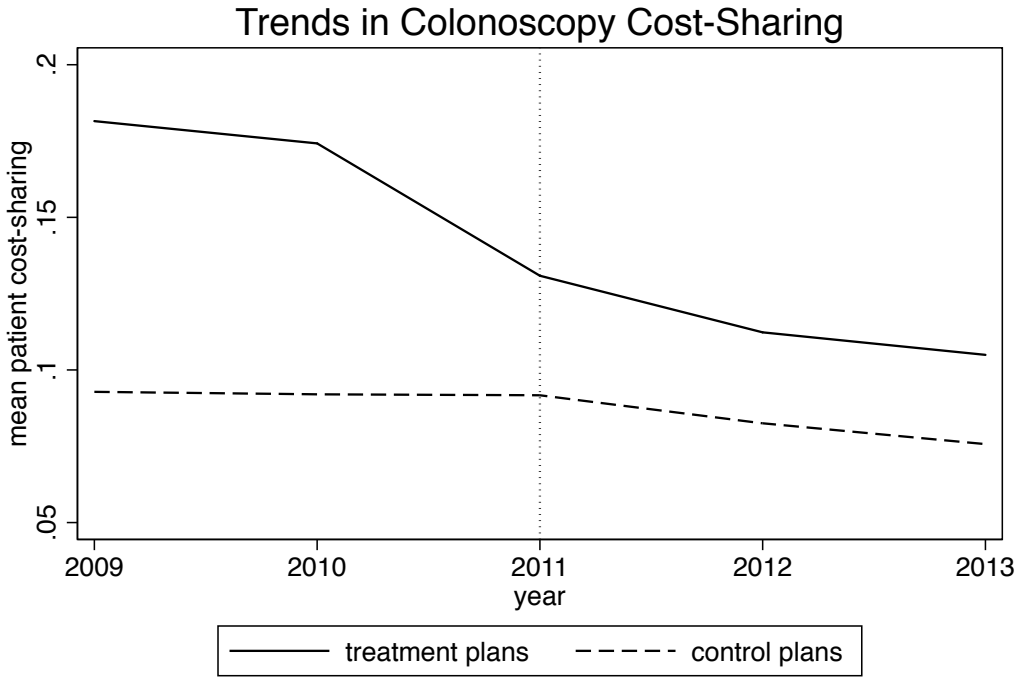


Table 1: Colonoscopy Utilization Descriptive Characteristics

	Pre-ACA Implementation (2009-2010)					Post-ACA Implementation (2011-2013)				
	Control Plan		Treatment Plan			Control Plan		Treatment Plan		
	Mean	SD	Mean	SD	Std. Diff	Mean	SD	Mean	SD	Std. Diff
Above age 60	41.7%	49.3%	43.3%	49.6%	-0.034	46.6%	49.9%	48.0%	50.0%	-0.029
% Male	51.9%	50.0%	53.4%	49.9%	-0.031	52.0%	50.0%	53.5%	49.9%	-0.030
Chronic condition diagnoses										
Charlson score	0.35	0.83	0.36	0.85	-0.012	0.38	0.90	0.40	0.92	-0.015
Acute myocardial infarction	1.1%	10.3%	1.1%	10.3%	-0.001	1.0%	10.1%	1.0%	10.1%	0.000
Congestive heart failure	0.8%	8.9%	0.8%	9.0%	-0.002	0.9%	9.3%	0.9%	9.6%	-0.005
Peripheral vascular disease	0.8%	9.1%	0.8%	9.0%	0.002	0.8%	8.9%	0.8%	9.0%	-0.002
Cerebrovascular disease	2.1%	14.5%	2.2%	14.6%	-0.004	2.2%	14.8%	2.3%	15.0%	-0.005
Dementia	0.0%	1.7%	0.0%	1.9%	-0.003	0.0%	1.9%	0.0%	2.0%	-0.003
COPD	7.5%	26.3%	7.5%	26.4%	-0.002	7.6%	26.5%	7.6%	26.5%	-0.001
Rheumatoid disease	1.3%	11.5%	1.4%	11.8%	-0.005	1.4%	11.9%	1.5%	12.1%	-0.004
Peptic ulcer disease	0.4%	6.4%	0.4%	6.4%	-0.001	0.4%	6.3%	0.4%	6.3%	-0.001
Mild liver disease	0.3%	5.3%	0.3%	5.4%	-0.003	0.3%	5.6%	0.3%	5.7%	-0.003
Diabetes	11.1%	31.4%	11.4%	31.7%	-0.010	11.8%	32.3%	12.2%	32.7%	-0.012
Diabetes + complications	1.6%	12.5%	1.7%	12.9%	-0.009	1.9%	13.6%	2.0%	13.9%	-0.007
Hemiplegia or paraplegia	0.1%	3.2%	0.1%	3.3%	-0.003	0.1%	3.5%	0.1%	3.7%	-0.003
Renal disease	1.1%	10.7%	1.2%	10.9%	-0.006	1.4%	11.9%	1.5%	12.3%	-0.008
Cancer	4.1%	19.8%	4.1%	19.9%	-0.002	4.5%	20.7%	4.5%	20.7%	-0.001
Severe liver disease	0.1%	3.0%	0.1%	3.1%	-0.003	0.1%	3.5%	0.1%	3.6%	-0.004
Metastatic cancer	0.3%	5.7%	0.4%	5.9%	-0.005	0.4%	6.7%	0.5%	6.9%	-0.004
HIV/AIDS	0.2%	4.8%	0.3%	5.1%	-0.007	0.3%	5.0%	0.3%	5.6%	-0.012
Insurance deductible										
High deductible plan	11.4%	31.8%	13.9%	34.5%	-0.073	21.9%	41.3%	22.6%	41.8%	-0.018
Annual Medical Spending										
Inpatient	$998	$8,919	$979	$9,041	0.002	$1,231	$10,815	$1,269	$13,643	-0.003
Outpatient	$1,489	$6,603	$1,497	$6,587	-0.001	$1,927	$9,516	$1,941	$9,522	-0.002
Office-based	$2,105	$5,352	$2,118	$5,321	-0.002	$2,410	$6,634	$2,428	$6,712	-0.003
Prescription drug	$689	$2,496	$686	$2,509	0.001	$726	$3,277	$726	$3,282	0.000
Total	$5,281	$15,737	$5,280	$15,734	0.000	$6,294	$20,088	$6,364	$22,005	-0.003
Medical Utilization										
Inpatient visits	0.1	0.4	0.1	0.4	0.007	0.1	0.5	0.1	0.5	0.002
Outpatient visits	1.4	3.0	1.4	2.9	0.017	1.5	3.5	1.5	3.5	0.016
Office-based visits	12.6	15.7	12.6	15.5	-0.005	13.4	16.8	13.5	16.9	-0.005
ED visits	0.1	0.3	0.1	0.4	-0.051	0.1	0.4	0.1	0.4	-0.046
Inpatient days	4.8	188.1	4.4	119.8	0.002	6.1	289.8	7.0	692.4	-0.002
Prescription drug fills	7.8	16.4	8.0	16.9	-0.015	7.7	16.7	7.9	17.1	-0.010
Colonoscopy Utilization										
Colonoscopy cost-sharing ($)	$132.2	$132.3	$219.4	$143.5	-0.631	$135.3	$148.5	$191.3	$122.4	-0.412
Colonoscopy cost-sharing (%)	12.0%	12.0%	21.3%	11.6%	-0.784	10.8%	11.5%	16.2%	9.1%	-0.520
Colonoscopy in year (%)	8.0%	27.1%	7.7%	26.6%	0.012	7.7%	26.7%	7.6%	26.5%	0.005
Observations	431,477		500,271			687,219		793,274		

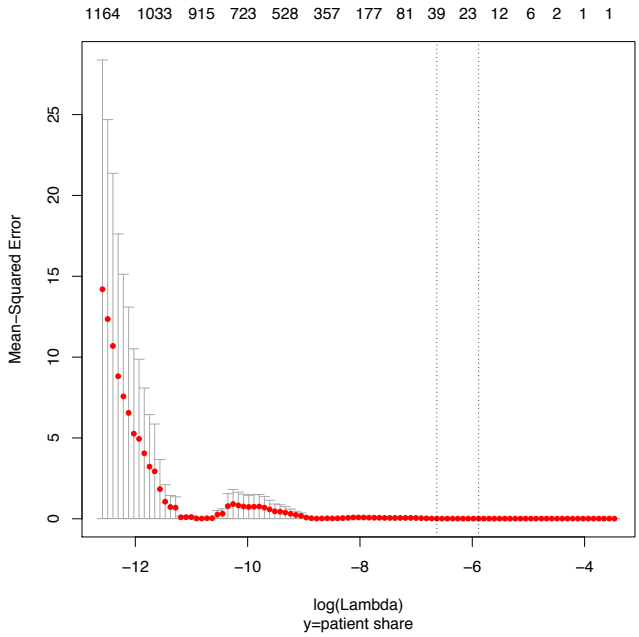
This table presents descriptive characteristics between the treatment and control plans, before and after the 2011 implementation of the ACA's cancer screening coverage requirements. For period, this table presents means for each population, standard deviations for each population, and standardized differences between the two populations.

Table 2: Colonoscopy Cost Descriptive Characteristics

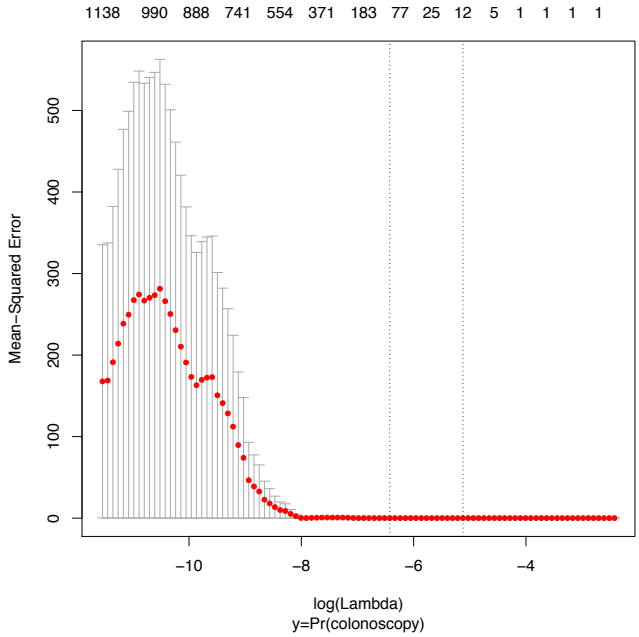
	Pre-ACA Implementation (2009-2010)					Post-ACA Implementation (2011-2013)				
	Control Plan		Treatment Plan			Control Plan		Treatment Plan		
	Mean	SD	Mean	SD	Std. Diff	Mean	SD	Mean	SD	Std. Diff
Above age 60	50.5%	50.0%	51.1%	50.0%	-0.011	51.6%	50.0%	51.9%	50.0%	-0.006
% Male	51.9%	50.0%	53.2%	49.9%	-0.026	52.2%	50.0%	53.4%	49.9%	-0.024
Chronic condition diagnoses										
Charlson score	0.56	1.13	0.56	1.14	0.002	0.54	1.12	0.55	1.14	-0.003
Acute myocardial infarction	1.5%	12.0%	1.4%	11.9%	0.003	1.3%	11.1%	1.2%	10.8%	0.006
Congestive heart failure	1.1%	10.5%	1.1%	10.4%	0.002	1.0%	9.9%	1.0%	9.8%	0.001
Peripheral vascular disease	1.2%	10.9%	1.2%	10.8%	0.003	1.0%	9.9%	1.0%	9.9%	0.001
Cerebrovascular disease	3.1%	17.4%	3.2%	17.5%	-0.002	2.9%	16.7%	2.9%	16.8%	-0.001
Dementia	0.1%	2.3%	0.1%	2.4%	-0.002	0.0%	2.1%	0.0%	2.1%	-0.001
COPD	10.1%	30.1%	9.7%	29.6%	0.013	9.6%	29.5%	9.3%	29.1%	0.010
Rheumatoid disease	1.7%	13.0%	1.7%	13.0%	0.001	1.8%	13.1%	1.8%	13.1%	-0.001
Peptic ulcer disease	1.6%	12.4%	1.6%	12.4%	0.001	1.4%	11.9%	1.4%	11.8%	0.003
Mild liver disease	0.7%	8.5%	0.7%	8.5%	-0.001	0.7%	8.4%	0.7%	8.3%	0.002
Diabetes	14.1%	34.8%	14.1%	34.8%	0.001	13.9%	34.6%	13.9%	34.6%	0.000
Diabetes + complications	2.2%	14.8%	2.2%	14.6%	0.004	2.3%	14.9%	2.3%	15.0%	-0.002
Hemiplegia or paraplegia	0.1%	3.4%	0.1%	3.5%	-0.001	0.1%	3.4%	0.1%	3.4%	0.000
Renal disease	1.8%	13.3%	1.8%	13.1%	0.003	1.9%	13.5%	1.9%	13.7%	-0.005
Cancer	8.1%	27.3%	8.1%	27.2%	0.003	7.8%	26.9%	7.9%	26.9%	-0.001
Severe liver disease	0.3%	5.6%	0.3%	5.6%	0.001	0.3%	5.6%	0.3%	5.6%	0.000
Metastatic cancer	0.9%	9.6%	1.0%	9.8%	-0.004	0.9%	9.5%	0.9%	9.7%	-0.004
HIV/AIDS	0.2%	5.0%	0.3%	5.1%	-0.002	0.3%	5.3%	0.3%	5.6%	-0.007
Insurance deductible										
High deductible plan	11.7%	32.1%	13.5%	34.2%	-0.055	21.4%	41.0%	23.5%	42.4%	-0.051
Annual Medical Spending										
Inpatient	$1,875	$12,574	$1,836	$13,355	0.003	$1,932	$13,117	$1,982	$13,876	-0.004
Outpatient	$3,821	$10,635	$3,805	$10,646	0.001	$4,368	$13,391	$4,379	$12,747	-0.001
Office-based	$4,231	$7,805	$4,284	$7,935	-0.007	$4,494	$8,249	$4,586	$8,770	-0.011
Prescription drug	$951	$2,857	$972	$2,925	-0.007	$960	$3,654	$938	$3,789	0.006
Total	$10,879	$23,126	$10,898	$23,496	-0.001	$11,754	$25,537	$11,885	$26,130	-0.005
Medical Utilization										
Inpatient visits	0.1	0.6	0.1	0.6	0.007	0.1	0.6	0.1	0.5	0.004
Outpatient visits	3.3	4.2	3.2	4.2	0.024	3.3	4.4	3.2	4.4	0.022
Office-based visits	21.0	19.7	20.9	19.5	0.002	21.2	19.5	21.2	19.6	0.001
ED visits	0.1	0.5	0.1	0.5	-0.051	0.1	0.4	0.1	0.6	-0.052
Inpatient days	9.3	278.7	9.5	301.9	-0.001	8.4	192.1	8.5	186.4	-0.001
Prescription drug fills	10.7	19.5	11.3	20.1	-0.026	10.2	19.2	10.0	19.1	0.007
Colonoscopy Costs and Cost-Sharing										
Colonoscopy cost ($)	$1,266.0	$896.4	$1,239.0	$887.0	0.031	$1,420.0	$1,040.0	$1,395.0	$1,042.7	0.024
Patient cost-sharing ($)	$111.6	$294.3	$200.8	$339.6	-0.281	$111.2	$320.9	$150.9	$343.7	-0.119
Patient cost-sharing (%)	9.2%	21.3%	17.8%	25.2%	-0.366	8.4%	21.2%	11.7%	22.9%	-0.148
Observations	537,295		492,897			743,037		841,839		

This table presents descriptive characteristics between the treatment and control plans, before and after the 2011 implementation of the ACA's cancer screening coverage requirements. For period, this table presents means for each population, standard deviations for each population, and standardized differences between the two populations.

Figure 2: Colonoscopy Utilization Covariate Selection
(a) First Stage Variable Selection

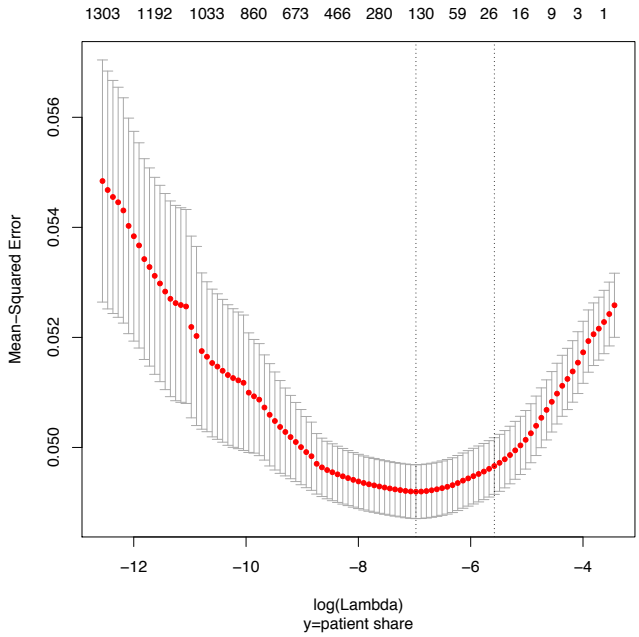


(b) Reduced Form Variable Selection

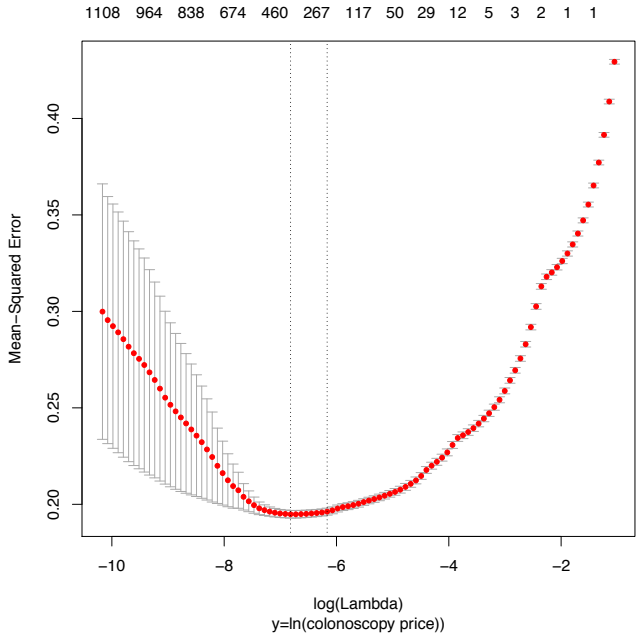


These figures plot the cross-validated mean-squared error curves for each lambda. The top panel uses plan-level patient cost-sharing as the dependent variable, while the bottom panel uses the probability of receiving a colonoscopy as the dependent variable. In each figure, the left vertical line indicates the λ penalization parameter value that has the minimum mean cross-validated error. The right vertical line indicates the λ parameter value within one standard deviation of the minimum mean cross-validated error.

Figure 3: Colonoscopy Costs Covariate Selection
(a) First Stage Variable Selection



(b) Reduced Form Variable Selection



These figures plot the cross-validated mean-squared error curves for each lambda. The top panel uses procedure-level patient cost-sharing as the dependent variable, while the bottom panel uses the log-transformed colonoscopy price as the dependent variable. In each figure, the left vertical line indicates the λ penalization parameter value that has the minimum mean cross-validated error. The right vertical line indicates the λ parameter value within one standard deviation of the minimum mean cross-validated error.

Table 3: First Stage Results: Effect of Exposure to ACA on Colonoscopy Cost-Sharing

	(1)	(2)	(3)	(4)
	Pre-selected covariates		*Algorithm-selected covariates*	
Panel A: Plan-level cost-sharing				
ACA x treatment plan	-0.0387***	-0.0387***	-0.0387***	-0.0388***
	(0.00525)	(0.00490)	(0.00531)	(0.00486)
treatment plan	0.0910***		0.0871***	
	(0.0101)		(0.0100)	
ACA	-0.0154***	-0.00946***	-0.0150***	-0.00936***
	(0.00362)	(0.00289)	(0.00368)	(0.00289)
Plan fixed effects		X		X
Observations	2,412,241	2,412,241	2,412,241	2,412,241
R-squared	0.271	0.796	0.229	0.797
Pre-ACA mean for treatment plans	0.213	0.213	0.213	0.213
Panel B: Procedure-level cost-sharing				
ACA x treatment plan	-0.0529***	-0.0523***	-0.0510***	-0.0515***
	(0.00452)	(0.00476)	(0.00449)	(0.00479)
treatment plan	0.0801***		0.0754***	
	(0.00625)		(0.00658)	
ACA	-0.0115***	-0.00904***	-0.00825***	-0.00593***
	(0.00215)	(0.00232)	(0.00218)	(0.00230)
Plan fixed effects		X		X
Observations	2,715,068	2,715,068	2,715,068	2,715,068
R-squared	0.079	0.185	0.092	0.208
Pre-ACA mean for treatment plans	0.178	0.178	0.178	0.178

This table presents first stage regression results that examine the effect of the instrument, exposure to the ACA's cancer screening coverage requirement, on cost-sharing for colonoscopies. Columns 1-2 include the pre-specified covariates and columns 3-4 include the algorithm-selected covariates. Columns 2 and 4 include insurance plan fixed effects. Panel A presents results for plan-level cost-sharing, while Panel B restricts the sample to procedure-level cost-sharing. Robust standard errors clustered at the insurance plan level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 4: Reduced Form Results: Effect of Exposure to ACA on Colonoscopy Utilization and Price Shopping

	(1)	(2)	(3)	(4)
	Pre-selected covariates		*Algorithm-selected covariates*	
Panel A: Utilization				
ACA x treatment plan	0.00224**	0.00228**	0.00247***	0.00224***
	(0.000930)	(0.000887)	(0.000794)	(0.000802)
treatment plan	-0.00262**		-0.00234**	
	(0.00106)		(0.00101)	
ACA	-0.00619***	-0.00554***	-0.00781***	-0.00717***
	(0.000639)	(0.000642)	(0.000593)	(0.000584)
Plan fixed effects		X		X
Observations	2,412,241	2,412,241	2,412,241	2,412,241
R-squared	0.038	0.046	0.139	0.147
Pre-ACA mean for treatment plans	0.077	0.077	0.077	0.077
Panel B: Price Shopping				
ACA x treatment plan	0.00147	0.00478	0.000789	0.00193
	(0.00572)	(0.00501)	(0.00246)	(0.00231)
treatment plan	-0.0206*		-0.0114**	
	(0.0108)		(0.00447)	
ACA	0.0884***	0.0861***	0.0375***	0.0390***
	(0.00368)	(0.00291)	(0.00182)	(0.00164)
Plan fixed effects		X		X
Observations	2,715,068	2,715,068	2,715,068	2,715,068
R-squared	0.038	0.143	0.555	0.580
Pre-ACA mean for treatment plans	$1,238.6	$1,238.6	$1,238.6	$1,238.6

This table presents reduced form regression results that examine the effect of the instrument, exposure to the ACA's cancer screening coverage requirement, on colonoscopy utilization and price shopping. Columns 1-2 include the pre-specified covariates and columns 3-4 include the algorithm-selected covariates. Columns 2 and 4 include insurance plan fixed effects. Panel A presents results for colonoscopy utilization, while Panel B presents results for colonoscopy price-shopping. Robust standard errors clustered at the insurance plan level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 5: OLS Results: Association Between Cost-Sharing and Colonoscopy Utilization and Price Shopping

	(1)	(3)	(3)	(4)
	Pre-selected covariates		*Algorithm-selected covariates*	
Panel A: Utilization				
plan-level cost sharing	-0.00948**	-0.0243***	0.0119***	-0.0186***
	(0.00404)	(0.00335)	(0.00307)	(0.00308)
Plan fixed effects		X		X
Observations	2,412,241	2,412,241	2,412,241	2,412,241
R-squared	0.038	0.046	0.139	0.147
Pre-ACA mean for treatment plans	0.077	0.077	0.077	0.077
Panel B: Price Shopping				
procedure-level cost sharing	-0.198***	-0.168***	-0.0664***	-0.0565***
	(0.0124)	(0.0114)	(0.00479)	(0.00534)
Plan fixed effects		X		X
Observations	2,715,068	2,715,068	2,715,068	2,715,068
R-squared	0.044	0.147	0.555	0.581
Pre-ACA mean for treatment plans	$1,238.6	$1,238.6	$1,238.6	$1,238.6

This table presents OLS regression results that examine the association between patient cost sharing and colonoscopy utilization and price shopping. Columns 1-2 include the pre-specified covariates and columns 3-4 include the algorithm-selected covariates. Columns 2 and 4 include insurance plan fixed effects. Panel A presents results for colonoscopy utilization, while Panel B presents results for colonoscopy price-shopping. Robust standard errors clustered at the insurance plan level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 6: 2SLS Results: Effect of Cost-Sharing on Colonoscopy Utilization and Price Shopping

	(1)	(2)	(3)	(4)	(5)	(6)
	No covariates		*Pre-selected covariates*		*Selected covariates*	
Panel A: Utilization						
plan-level cost sharing	-0.0533**	-0.0535*	-0.0580**	-0.0492**	-0.0642**	-0.0581***
	(0.0218)	(0.0219)	(0.0227)	(0.0217)	(0.0206)	(0.0202)
Plan fixed effects		X		X		X
Observations	2,412,241	2,412,241	2,412,241	2,412,241	2,412,241	2,412,241
R-squared	0.000	0.000	0.0238	0.036	0.138	0.138
Cragg-Donald F-stat	18,230	7,187	21,977	7,790	20,866	78,117
Kleibergen-Paap F-stat	50.07	62.27	54.69	63.03	54.06	64.28
Pre-ACA mean for treatment plans	0.077	0.077	0.077	0.077	0.077	0.077
Panel B: Price Shopping						
procedure-level cost-sharing	0.00589	-0.0537	-0.080	0.0663	-0.00386	-0.0266
	(0.121)	(0.114)	(0.106)	(0.0910)	(0.0481)	(0.0439)
Plan fixed effects		X		X		X
Observations	2,715,068	2,715,068	2,715,068	2,715,068	2,715,068	2,715,068
R-squared	0.007	0.009	0.040	0.031	0.5551	0.524
Cragg-Donald F-stat	8,788	9,760	9,365	9,781	8,764	9,758
Kleibergen-Paap F-stat	116	116.4	136.4	122.8	125.05	117.4
Pre-ACA mean for treatment plans	$1238.6	$1238.6	$1238.6	$1238.6	$1238.6	$1238.6

This table presents 2SLS regression results that examine the effect of patient cost sharing on colonoscopy utilization and price shopping. Columns 1-2 do not include covariates, columns 3-4 include the pre-specified covariates, and columns 5-6 include the algorithm-selected covariates. Columns 2, 4, and 6 include insurance plan fixed effects. Panel A presents results for colonoscopy utilization, while Panel B presents results for colonoscopy price-shopping. Robust standard errors clustered at the insurance plan level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

SUPPLEMENTAL APPENDIX

1. ROBUSTNESS TESTS

1.1. Continuous Cost-Sharing as an Instrument

The identification strategy of the main results separates insurance plans into those impacted by the ACA's cost-sharing requirement and the plans that were unaffected by the policy. As an additional approach, I use the continuous change in each plan's mean colonoscopy cost-sharing before and after the ACA as an instrumental variable.

In general, the 2SLS results are similar to the main results. As shown in top panel of Table A1.1, the colonoscopy utilization estimates are smaller than in the main results. The preferred results in column 4 indicate that a 10-percentage point decrease in colonoscopy cost-sharing leads to a 0.22 percentage point increase in the probability of having a colonoscopy. The bottom panel shows that the price-margin estimates are not statistically significant, and are smaller in magnitude. As in the main results, the predictive power of the double-selection models is much higher than in the pre-selected models.

The first-stage results on the effect of the ACA on cost-sharing have a slightly different interpretation than in the main results. For example, column 4 of Panel A in Table A1.2, which shows the first-stage results for the utilization analysis, indicates that the ACA lead to a 0.37 percentage point reduction in cost-sharing for every percentage point in the plans mean cost-sharing prior to the ACA. In other words, the ACA reduced within-plan cost-sharing by slightly less than 40%. The bottom panel presents similar results for the price-margin. Column 4 indicates that the ACA reduced within-plan cost-sharing by approximately 45%. Not

surprisingly, in both tables, the coefficients in columns 1 and 3 on the pre-ACA mean cost-sharing variable are close to 1.

The reduced form results also have a slightly different interpretation. The results in Panel A of Table A1.3 show that the ACA led to an approximately 1-percentage point increase in the use of colonoscopies for every 1-percentage point reduction in cost-sharing. The reduced form results in Panel B of Table A1.3 do not show any intensive-margin response to the ACA based on differences in pre-ACA cost-sharing.

Appendix Table A1.1: Effect of Cost-Sharing on Colonoscopy Utilization and Price Shopping Using Continuous Exposure to ACA

	(1)	(2)	(3)	(4)
	Pre-selected covariates		*Algorithm-selected covariates*	
Panel A: Utilization				
plan-level cost sharing	-0.0229***	-0.0309***	-0.0225***	-0.0284***
	(0.00821)	(0.00887)	(0.00794)	(0.00858)
Plan fixed effects		X		X
Observations	2,412,241	2,411,835	2,412,241	2,411,835
R-squared	0.038	0.036	0.139	0.138
Cragg-Donald F-stat	187324	392603	188296	392713
Kleibergen-Paap F-stat	189.8	162.8	187	163.1
Panel B: Price Shopping				
procedure-level cost-sharing	0.0269	0.0267	-0.00875	-0.00351
	(0.0408)	(0.0321)	(0.0189)	(0.0188)
Plan fixed effects		X		X
Observations	2,715,068	2,714,501	2,715,068	2,714,501
R-squared	0.040	0.028	0.555	0.524
Cragg-Donald F-stat	49926	39254	51056	39778
Kleibergen-Paap F-stat	570	384.5	565.5	374.3

This table presents 2SLS regression results that examine the effect of patient cost sharing on colonoscopy utilization. Plan-level cost-sharing is instrumented for by the each plan's mean colonoscopy cost-sharing in the pre-ACA period. Columns 1-2 include the pre-specified covariates and columns 3-4 include the algorithm-selected covariates. Columns 2 and 4 include insurance plan fixed effects. Panel A presents results for colonoscopy utilization, while Panel B presents results for colonoscopy price-shopping. Robust standard errors clustered at the insurance plan level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table A1.2: First Stage Results: Effect of Continuous Exposure to ACA on Colonoscopy Cost-Sharing

	(1)	(2)	(3)	(4)
	Pre-selected covariates		*Algorithm-selected covariates*	
Panel A: Plan-level cost-sharing				
ACA x pre-ACA cost-sharing	-0.386***	-0.370***	-0.386***	-0.370***
	(0.0280)	(0.0291)	(0.0283)	(0.0291)
pre-ACA cost-sharing	0.962***		0.969***	
	(0.0250)		(0.0221)	
2010	-0.00347*	-0.00189	-0.00329	-0.00187
	(0.00208)	(0.00210)	(0.00207)	(0.00210)
2011	0.0308***	0.0310***	0.0309***	0.0311***
	(0.00395)	(0.00417)	(0.00392)	(0.00417)
2012	0.0175***	0.0192***	0.0176***	0.0193***
	(0.00388)	(0.00411)	(0.00386)	(0.00411)
2013	0.0107***	0.0136***	0.0108***	0.0137***
	(0.00394)	(0.00422)	(0.00394)	(0.00421)
Plan fixed effects		X		X
Observations	2,412,241	2,412,241	2,412,241	2,412,241
R-squared	0.579	0.822	0.580	0.822
Panel B: Procedure-level cost-sharing				
ACA x pre-ACA cost-sharing	-0.485***	-0.444***	-0.484***	-0.440***
	(0.0203)	(0.0227)	(0.0204)	(0.0229)
pre-ACA cost-sharing	0.926***		0.931***	
	(0.00700)		(0.00792)	
2010	-0.00683***	-0.00673***	-0.00642***	-0.00663***
	(0.00164)	(0.00187)	(0.00169)	(0.00191)
2011	0.0401***	0.0343***	0.0424***	0.0360***
	(0.00305)	(0.00351)	(0.00312)	(0.00352)
2012	0.0263***	0.0202***	0.0302***	0.0236***
	(0.00301)	(0.00353)	(0.00308)	(0.00353)
2013	0.0210***	0.0152***	0.0261***	0.0198***
	(0.00352)	(0.00417)	(0.00360)	(0.00411)
Plan fixed effects		X		X
Observations	1,922,738	1,922,738	1,922,738	1,922,738
R-squared	0.165	0.203	0.173	0.212

This table presents first stage regression results that examine the effect of the alternative instrument, pre-ACA plan-level cost-sharing interacted with the implementation of the ACA's cancer screening coverage requirement, on cost-sharing for colonoscopies. Columns 1-2 include the pre-specified covariates and columns 3-4 include the algorithm-selected covariates. Columns 2 and 4 include insurance plan fixed effects. Panel A presents results for plan-level cost-sharing, while Panel B restricts the sample to procedure-level cost-sharing. Robust standard errors clustered at the insurance plan level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table A1.3: Reduced Form Results: Effect of Continuous Exposure to ACA on Colonoscopy Utilization

	(1)	(2)	(3)	(4)
	Pre-selected covariates		*Algorithm-selected covariates*	
Panel A: Utilization				
ACA x pre-ACA cost-sharing	0.00883***	0.0114***	0.00868***	0.0105***
	(0.00329)	(0.00342)	(0.00314)	(0.00325)
pre-ACA cost-sharing	-0.0135***		0.0136***	
	(0.00452)		(0.00414)	
2010	-0.00289***	-0.00267***	-0.00369***	-0.00348***
	(0.000636)	(0.000629)	(0.000548)	(0.000554)
2011	-0.00626***	-0.00595***	-0.00770***	-0.00750***
	(0.000868)	(0.000859)	(0.000752)	(0.000760)
2012	-0.00843***	-0.00801***	-0.0101***	-0.00982***
	(0.000853)	(0.000831)	(0.000748)	(0.000748)
2013	-0.00856***	-0.00805***	-0.0109***	-0.0106***
	(0.000886)	(0.000887)	(0.000777)	(0.000787)
Plan fixed effects		X		X
Observations	2,412,241	2,412,241	2,412,241	2,412,241
R-squared	0.038	0.046	0.139	0.147
Panel B: Price Shopping				
ACA x pre-ACA cost-sharing	-0.0130	-0.0118	0.00424	0.00155
	(0.0198)	(0.0144)	(0.00911)	(0.00831)
pre-ACA cost-sharing	-0.259***		-0.131***	
	(0.0449)		(0.0154)	
2010	0.0327***	0.0340***	0.0102***	0.0113***
	(0.00264)	(0.00216)	(0.00139)	(0.00130)
2011	0.0727***	0.0760***	0.0273***	0.0308***
	(0.00426)	(0.00337)	(0.00210)	(0.00184)
2012	0.110***	0.111***	0.0429***	0.0469***
	(0.00512)	(0.00400)	(0.00241)	(0.00207)
2013	0.145***	0.147***	0.0597***	0.0646***
	(0.00575)	(0.00485)	(0.00270)	(0.00254)
Plan fixed effects		X		X
Observations	2,715,068	2,715,068	2,715,068	2,715,068
R-squared	0.041	0.144	0.555	0.580

This table presents reduced form regression results that examine the effect of the alternative instrument, pre-ACA plan-level cost-sharing interacted with the implementation of the ACA's cancer screening coverage requirement, on colonoscopy utilization. Columns 1-2 include the pre- specified covariates and columns 3-4 include the algorithm-selected covariates. Columns 2 and 4 include insurance plan fixed effects. Panel A presents results for colonoscopy utilization, while Panel B presents results for colonoscopy price-shopping. Robust standard errors clustered at the insurance plan level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

1.2. Pre-ACA Trend Differences Between Treatment and Control Plans

The instrumental variables approach used in this paper relies on the differential exogenous shock of the ACA's requirements to the treatment plans relative to the control plans. To test the exogeneity of the ACA's impact, I test for pre-trend differences between the treatment and control plans. To do so, I estimate the same 2SLS regression as in the main results, but replace the $\boldsymbol{ACA_t \times treatment_{ij}}$ variable with interactions of the treatment plans and the years 2010, 2011, 2012, and 2013. 2010 is the year prior to the ACA's implementation, and so the $2010_t \times treatment_{ij}$ coefficient gives the pre-ACA trend difference between the treatment and control plans.

As shown in Panel A of Appendix Table A2.1 there is no pre-trend in the first stage results for plan-level cost-sharing. Panel B shows a small pre-trend in columns 1 and 3 for procedure-level cost-sharing, but this effect disappears once plan fixed effects are included (columns 2 and 4). Appendix Table A2.2 presents similar results, but focuses on the reduced form results. For both dependent variables, the $2010_t \times treatment_{ij}$ coefficients are not distinguishable from zero.

Appendix Table A2.1: First Stage Instrument Pre-Trends Test

	(1)	(2)	(3)	(4)
	Pre-selected covariates		*Algorithm-selected covariates*	
Panel A: Plan-level cost-sharing				
treatment plan x 2010	-0.00493	-0.00468	-0.00473	-0.00482
	(0.00420)	(0.00423)	(0.00403)	(0.00424)
treatment plan x 2011	-0.0354***	-0.0351***	-0.0353***	-0.0352***
	(0.00493)	(0.00495)	(0.00498)	(0.00494)
treatment plan x 2012	-0.0454***	-0.0451***	-0.0452***	-0.0453***
	(0.00598)	(0.00561)	(0.00607)	(0.00560)
treatment plan x 2013	-0.0429***	-0.0430***	-0.0427***	-0.0432***
	(0.00771)	(0.00691)	(0.00789)	(0.00688)
Plan fixed effects		X		X
Observations	2,412,241	2,412,241	2,412,241	2,412,241
R-squared	0.275	0.799	0.233	0.799
Panel B: Procedure-level cost-sharing				
treatment plan x 2010	-0.00611*	-0.00514	-0.00610*	-0.00531
	(0.00363)	(0.00358)	(0.00359)	(0.00362)
treatment plan x 2011	-0.0506***	-0.0497***	-0.0492***	-0.0492***
	(0.00504)	(0.00519)	(0.00504)	(0.00522)
treatment plan x 2012	-0.0584***	-0.0578***	-0.0564***	-0.0570***
	(0.00543)	(0.00568)	(0.00537)	(0.00577)
treatment plan x 2013	-0.0586***	-0.0574***	-0.0567***	-0.0566***
	(0.00687)	(0.00748)	(0.00687)	(0.00757)
Plan fixed effects		X		X
Observations	2,715,068	2,715,068	2,715,068	2,715,068
R-squared	0.080	0.186	0.093	0.209

This table presents the first stage regression results that examine time trends in the effect of the instrument, exposure to the ACA's cancer screening coverage requirement, on colonoscopy utilization. Columns 1-2 include the pre-specified covariates and columns 3-4 include the algorithm-selected covariates. Columns 2 and 4 include insurance plan fixed effects. Panel A presents results for all colonoscopies, while Panel B restricts the sample to colonoscopies performed for screening purposes. Robust standard errors clustered at the insurance plan level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix Table A2.2: Reduced Form Instrument Pre-Trends Test

	(1)	(2)	(3)	(4)
	Pre-selected covariates		*Algorithm-selected covariates*	
Panel A: Utilization				
treatment plan x 2010	0.00003	0.000149	0.000806	0.000605
	(0.00129)	(0.00127)	(0.00112)	(0.00113)
treatment plan x 2011	0.00129	0.00135	0.00245**	0.00217*
	(0.00146)	(0.00139)	(0.00119)	(0.00119)
treatment plan x 2012	0.00323**	0.00338**	0.00332***	0.00301**
	(0.00147)	(0.00141)	(0.00123)	(0.00125)
treatment plan x 2013	0.00225	0.00236*	0.00288**	0.00250**
	(0.00148)	(0.00141)	(0.00123)	(0.00125)
Plan fixed effects		X		X
Observations	2,412,241	2,412,241	2,412,241	2,412,241
R-squared	0.038	0.046	0.139	0.147
Panel B: Price Shopping				
treatment plan x 2010	0.00431	0.00262	0.000581	-0.000530
	(0.00533)	(0.00415)	(0.00275)	(0.00249)
treatment plan x 2011	0.00276	0.00265	0.00112	0.000638
	(0.00621)	(0.00491)	(0.00301)	(0.00274)
treatment plan x 2012	0.00556	0.00793	0.00150	0.00203
	(0.00798)	(0.00635)	(0.00351)	(0.00321)
treatment plan x 2013	-0.000696	0.00406	-0.000837	0.000612
	(0.00955)	(0.00831)	(0.00427)	(0.00414)
Plan fixed effects		X		X
Observations	2,715,068	2,715,068	2,715,068	2,715,068
R-squared	0.039	0.144	0.555	0.580

This table presents reduced form regression results that examine time trends in the effect of the instrument, exposure to the ACA's cancer screening coverage requirement, on colonoscopy utilization. Columns 1-2 include the pre-specified covariates and columns 3-4 include the algorithm-selected covariates. Columns 2 and 4 include insurance plan fixed effects. Panel A presents results for colonoscopy utilization, while Panel B presents results for colonoscopy price-shopping. Robust standard errors clustered at the insurance plan level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

1.3. Colonoscopies Performed for Screening Purposes

As a sensitivity test, I also distinguished procedures performed solely for diagnostic or screening purposes from those accompanied with a therapeutic intervention. Colonoscopies performed for diagnostic purposes and not for routine screening are identified using ICD-9 codes for gastrointestinal symptoms and diseases reported within 6 months prior to colonoscopy (Lurie and Welch 1999). These codes included clostridium difficile colitis (8.45), ulcerative colitis (556), Crohn's disease (555), ischemic colitis (557), colitis NOS (558), anorectal bleeding (569.3), melena (578.1), gastrointestinal bleeding (286.5, 459.0, 562.02, 562.03, 562.12, 562.13, 562.3, 569.84-569.86, 678.9, 792.1, and 998.1) abdominal pain (789.00-789.09), abdominal swelling (789.09-789.30), abdominal tenderness (789.60-780.69), abdominal bloating (787.3), megacolon (564.7), change in bowel habits (564.0, 787.99, 787.91), diverticulitis or diverticular hemorrhage (562.11-562.13), volvulus (560.2), and a history of colorectal cancer (153-154.8, 239.0, 230.3-230.6. To identify elective diagnostic procedures performed to evaluate an abnormal test result, this study uses ICD-9 codes for iron deficiency anemia (280.0-280.9, 285.9) and abnormal stool contents (792.1).

Appendix Figure A3.1 presents the descriptive trends in patient cost-sharing for colonoscopies performed for screening purposes between the treatment and control plans. Similar to Figure 1, the treatment plans show a steep decrease in patient cost-sharing following the 2011 implementation of the ACA. Appendix Table A3.1 presents the 2SLS results that exclude colonoscopies performed only for diagnostic purposes. Overall the results are similar to the main results. As shown in the top panel, every 10-percentage point increase in cost-sharing leads to a 0.4 percentage point reduction in screening colonoscopy utilization. However, because the baseline of screening colonoscopies is lower than the utilization rate of all colonoscopies, 5.4

% compared to 7.7%, the estimated elasticity is nearly identical, 0.156. As shown in the bottom panel, there is still no intensive-margin effect on price-shopping.

Figure A3.1: Trends in Screening Colonoscopy Cost-Sharing by Insurance Plan Type

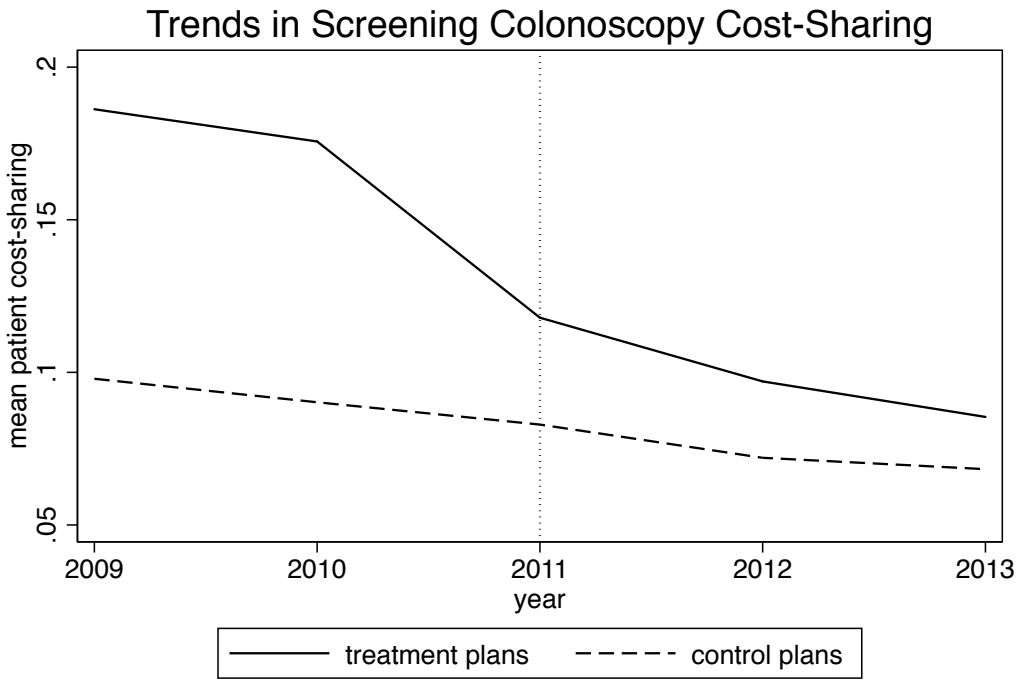


Table A3.1: Effect of Cost-Sharing on Screening Colonoscopy Utilization and Price Shopping

	(1)	(2)	(3)	(4)	(5)	(6)
	No covariates		*Pre-selected covariates*		*Selected covariates*	
Panel A: Utilization						
plan-level cost sharing	-0.0364*	-0.0359*	-0.0422**	-0.0408**	-0.0474***	-0.0402**
	(0.0191)	(0.0192)	(0.0189)	(0.0192)	(0.0179)	(0.0177)
Plan fixed effects		X		X		X
Observations	2,358,236	2,358,236	2,358,236	2,358,236	2,358,236	2,358,236
R-squared	0.000	0.000	0.022	0.002	0.1088	0.1083
Cragg-Donald F-stat	17,835	76,167	21,532	75,783	20,439	76,097
Kleibergen-Paap F-stat	50.17	62.11	54.52	62.86	54.16	64.09
Pre-ACA mean for treatment plans	0.054	0.054	0.054	0.054	0.054	0.054
Panel B: Price Shopping						
procedure-level cost-sharing	-0.0296	-0.0685	-0.0416	-0.0827	-0.003899	-0.02254
	(0.108)	(0.0989)	(0.0983)	(0.0816)	(0.0422)	(0.0382)
Plan fixed effects		X		X		X
Observations	1,922,738	1,922,738	1,922,738	1,922,738	1,922,738	1,922,738
R-squared	0.009	0.009	0.041	0.032	0.5948	0.563
Cragg-Donald F-stat	8,007	8,878	8,536	8,860	8,036	8,890
Kleibergen-Paap F-stat	142.7	144.1	158.9	143.6	148.51	141.3
Pre-ACA mean for treatment plans	$1213.5	$1213.5	$1213.5	$1213.5	$1213.5	$1213.5

This table presents robustness 2SLS regression results that examine the effect of patient cost sharing on screening colonoscopy utilization and price shopping. Columns 1-2 do not include covariates, columns 3-4 include the pre-specified covariates, and columns 5-6 include the algorithm-selected covariates. Columns 2, 4, and 6 include insurance plan fixed effects. Panel A presents results for colonoscopy utilization, while Panel B presents results for colonoscopy price-shopping. Robust standard errors clustered at the insurance plan level in parentheses. *** p<0.01, ** p<0.05, * p<0.1.