

# Cancer Screening Outreach Guided by Machine Learning: The Benefits of Proactive Care

Minje Park

Faculty of Business and Economics, The University of Hong Kong, Pokfulam Road, Hong Kong, parkmj@hku.hk

Carri W. Chan

Decision, Risk, and Operations, Columbia Business School, New York, NY 10027, cwchan@columbia.edu

Keith Boell

Geisinger, kaboell@geisinger.edu

Elliot G. Mitchell

Geisinger, emitchell3@geisinger.edu

Abdul A. Tariq

Children's Hospital of Philadelphia, tariqa3@chop.edu

David K. Vawdrey

Geisinger, dvawdrey@geisinger.edu

**Problem definition:** With the advance of data analytics, many disease prediction models have been developed with the intent of detecting diseases earlier and improving patient outcomes through earlier treatment. The operationalization of interventions and care based on these predictive models is critical to attaining these goals. We study the real-world effects of a machine learning-guided colorectal cancer screening outreach program deployed at a health system in Pennsylvania. **Methodology/results:** Using a regression discontinuity design based on the predicted risk score for having cancer, we find that the program increases the likelihood of colonoscopy uptake in three and six months by 6.0 percentage points (214% increase relative to the control sample within the bandwidth) and 6.9 percentage points (117% increase), respectively. Importantly, we also find significant effects on mortality. We estimate that the program decreases 2-year mortality by 6.2 percentage points (43% decrease). **Managerial implications:** Our finding suggests that a proactive cancer screening outreach program where individuals are selected for intervention based on a machine learning algorithm could significantly improve patient outcomes in addition to achieving higher disease detection rates. Our analysis demonstrates an analytical framework for rigorously evaluating machine learning-aided outreach programs for other cancers and diseases. Establishing unbiased estimates of the impact of machine learning-aided screening outreach is critical for capacity planning of screening resources, such as colonoscopies.

**Key words:** Machine learning, colorectal cancer, screening, mortality, regression discontinuity

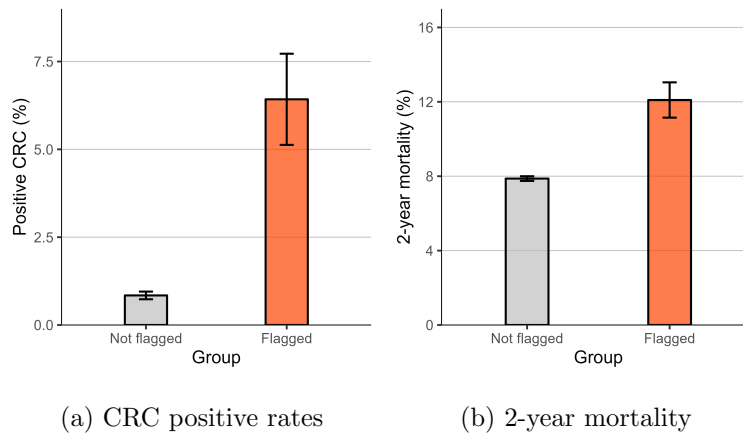
---

The authors thank the Department Editor, Jérémie Gallien, the Associate Editor, the reviewers, and the 2025 M&SOM Practice-Based Research Competition Industry Judge Panel – Jillian Berry Jaeker, Nicolas Stier, and Théophane Weber – for their constructive comments and suggestions. The authors acknowledge Jim Urick for excellent technical and deployment support for the CRC screening program at Geisinger, and Rebecca Maff for project management. We thank the Care Gaps Nurses team at Geisinger, who handle direct patient outreach. The authors are also grateful to Medial EarlySign, Ming Hu, and Michael Lingzhi Li for valuable discussions.

## 1. Introduction

Colorectal cancer (CRC) is the second leading cause of cancer deaths in the U.S., causing approximately 52,000 deaths in the year 2023. Patients diagnosed with early-stage CRC have a 90.9% 5-year survival rate, while those with late-stage CRC have only a 15.6% survival rate ([National Cancer Institute 2024](#)). Unfortunately, only 35% of patients are diagnosed with CRC when it is still in an early, localized stage ([National Cancer Institute 2024](#)). This could be due to low compliance with screening recommendations. The U.S. Preventive Services Task Force recommends that all adults with an average risk of CRC start with regular screening at 45, which was lowered from 50 starting in 2021 ([U.S. Preventive Services Task Force 2021](#)). Nonetheless, 46% of U.S. adults aged 45 years and older were overdue for CRC screening (colonoscopy) in 2021 ([American Cancer Society 2023](#)). This is well below the goal of 80% CRC screening rates set by the American Cancer Society National Colorectal Cancer Roundtable ([American Cancer Society National Colorectal Cancer Roundtable 2024](#)). Although colonoscopies have high sensitivity and specificity for CRC detection and are effective in reducing mortality, participation has been low because of its invasiveness and discomfort ([Winawer et al. 1993](#), [Roy et al. 2006](#), [Zauber et al. 2012](#), [Shaukat and Levin 2022](#)).

In 2019, Geisinger – an integrated health system in Pennsylvania – implemented a program to engage high-risk patients who are overdue for CRC screenings by providing personalized outreach to advise and educate these individuals about the benefits of undergoing a colonoscopy and to invite them to schedule a colonoscopy. This program used a predictive machine learning (ML) algorithm to identify patients at higher risk for CRC among those who are noncompliant. While it is clear that patients who were flagged for intervention by the ML model due to high risk of CRC had significantly higher positive rates for CRC (flagged: 6.42%, not flagged: 0.84%, t-test  $p$ -value  $< 0.001$ ), they also had higher mortality rates (see [Figure 1](#)). It remains unclear whether a cancer screening policy guided by ML prediction would be successful when implemented in the field at a large health system, and if so, what the potential impact of this program would have. For example, for it to be successful, patients identified as high-risk must agree to undergo screening and treatments. We address this gap by studying the effects of an ML-aided cancer screening outreach program targeting CRC, which has been operationalized at Geisinger since 2019. We aim to understand whether targeted cancer screening outreach, which is guided by ML to select individuals for intervention, can impact patient behaviors and reduce mortality of lethal cancers such as CRC. Accurately estimating these effects is essential to 1) informing whether such outreach programs, guided by ML tools, can be helpful for providers screening for CRC and, perhaps more importantly, 2) understanding what, if any, resource allocations must be made to accommodate such a new program.

**Figure 1** Colonoscopy Screening Results and Mortality by Group

*Notes.* Figure shows colorectal cancer (CRC) positive rates and 2-year mortality by group (flagged vs. not flagged). The bar and error bar represent the mean value and the standard errors of the mean, respectively.

The effects of cancer screening outreach programs where interventions are targeted based on an ML algorithm are challenging for managers to evaluate because the treatment (i.e., being identified and flagged as high-risk patients) is a function of the risk of having cancer predicted by the ML. In other words, the treated units are not randomly selected into treatment. Therefore, individuals in the treated group (or flagged group identified as high-risk by the ML model) exhibit riskier characteristics regarding CRC and differ from those in the control group. For example, the treated group, on average, is older than the control group (flagged: 67.3, not flagged: 62.4, t-test  $p$ -value  $< 0.001$ ). A simple comparison could lead to a misleading conclusion. By comparing the 2-year mortality by group (flagged: 12.1%, not flagged: 7.90%, t-test  $p$ -value  $< 0.001$ ), one may naively conclude that the program doesn't reduce mortality (or even that it *increases* mortality). To overcome this challenge, we employ a regression discontinuity (RD) design by exploiting the fact that patients were flagged only if their CRC risk score exceeded a fixed threshold. We estimate the local treatment effect identified from the discontinuity in being flagged at the cutoff risk score.

Since its first introduction by [Thistlethwaite and Campbell \(1960\)](#), RD design has been used extensively in health economics and medical literature ([Calonico et al. 2024](#)). For example, [Card et al. \(2009\)](#) study the effects of Medicare coverage on the conditions of patients admitted via the emergency department by comparing patients across the age 65 threshold for Medicare eligibility. By applying the RD design, they find that a 20% reduction in deaths for admitted patients who are just over 65, relative to admitted patients just under 65, suggests the significant impact of Medicare coverage on patient survival. [Clark and Royer \(2013\)](#) analyze the causal effects of education on adult mortality and health by exploiting changes in British compulsory schooling laws as the discontinuity in educational attainment. Although the policy increased the length of education, no significant impact on health (mortality and self-reported health) or health behaviors (activity

level, smoking, and drinking) are found. Other examples in health economics applying RD designs include [Almond and Doyle Jr. \(2011\)](#), [Bharadwaj et al. \(2013\)](#), and [Adams et al. \(2022\)](#).

We use a nonparametric RD design with a local polynomial approach ([Imbens and Kalyanaraman 2012](#), [Cattaneo et al. 2020](#)) by approximating regression functions only near the cutoff instead of assuming a functional form of the regression of the entire set of observations. A key consideration in this RD design is bandwidth. Instead of relying on the researchers' discretion, we adopt a data-driven method to choose the bandwidth, which minimizes the mean squared error (MSE) of the local polynomial RD estimator (MSE-optimal bandwidth), developed in [Imbens and Kalyanaraman \(2012\)](#) and [Calonico et al. \(2014\)](#). The advantages of using MSE-optimal bandwidth are that it can be obtained in closed form, and the RD point estimator is consistent and MSE-optimal ([Cattaneo et al. 2020](#)).

Applying the RD design, we find that the ML-guided CRC screening outreach program increases the likelihood of colonoscopy uptake by 6.0 percentage points (pp) in 3 months (214% increase relative to the control group within the bandwidth) and 6.9 pp in 6 months (117% increase), and significantly reduces the time to have a colonoscopy by 124 days (39% decrease), both sharply at the cutoff risk score. This impact on colonoscopy participation is large, considering that about 46% of U.S. adults were overdue for colonoscopy in 2021 ([American Cancer Society 2023](#)). More importantly, we estimate that the program decreases 2-year mortality by 6.2 pp (43% reduction) at the cutoff, despite the fact that there was no discontinuity in the risk of having CRC at the cutoff. We provide potential explanations of the effects on mortality. Our results suggest that ML-guided screening outreach programs for lethal cancers like CRC could be impactful in increasing screening participation and decreasing mortality, in addition to higher detection rates. We verify our results with numerous robustness checks, including using a local quadratic regression, parametric RD, different bandwidths and kernel functions, placebo cutoffs, different samples, and log-transformed risk scores.

With the explosion of available data, ML has been adopted in many industries, including manufacturing, financial modeling, education, marketing, and healthcare ([Jordan and Mitchell 2015](#)). Especially in healthcare, disease prediction models have been developed widely by analyzing electronic medical records to estimate the probability of having a disease. Our work is one of the early studies that evaluate the impact of cancer screening outreach to individuals identified as high risk by an ML algorithm on patients' screening participation and mortality using field data. Our paper provides a framework to rigorously evaluate similar ML-aided outreach programs for other cancers and chronic diseases, a challenging task due to the non-randomized treatments. Estimating unbiased policy impact is critical for capacity planning of resource-intensive screenings, such as colonoscopy (the average cost of colonoscopy in the U.S. was \$2,125 ([Fisher et al. 2022](#))), as

over- or under-estimating the effects on screening participation could lead to significant waste or shortage of screening capacity. Our estimates suggest that by accurately estimating the treatment effect, hospitals could save \$40.4 per individual participating in the ML-guided screening program per year. Our paper guides managers in rigorously evaluating ML-guided policies, a challenging task due to the non-randomized treatments.

## 2. Related Literature

### 2.1. Interventions Using Algorithms in Healthcare

Our research complements research that develops and validates prediction algorithms for CRC by evaluating their real-world impact as part of hospital screening operations. [Kinar et al. \(2016, 2017\)](#) develop a model to identify individuals at increased risk for CRC by analyzing blood counts, age, and sex, and the model was validated in the U.K. ([Birks et al. 2017](#)), Northwestern U.S. ([Hornbrook et al. 2017](#)), and California ([Schneider et al. 2020](#)). For other types of cancer, [Gould et al. \(2021\)](#) develop a prediction model for lung cancer aiming to identify patients with early lung cancer, and [Stark et al. \(2019\)](#) develop ML models to predict five-year breast cancer risk using personal health data. Our work is one of the early evaluations of cancer screening programs guided by ML. We contribute to understanding the role of ML-guided CRC screening on patients by providing causal evidence of real-world policy effects.

Our research also builds on the literature that evaluates the interventions using algorithms in healthcare settings. [Bundorf et al. \(2024\)](#), using a randomized controlled trial, find that personalized information about insurance plans and digital expert advice changed consumers' insurance plan choices. [Adjerid et al. \(2023\)](#) study the effects of algorithm-enabled process innovation applied for sepsis care of hospitalized patients and find that the intervention reduces the likelihood of death from sepsis. [Kraus et al. \(2024\)](#) develop a decision model that allocates preventive care for diabetes mellitus type II based on ML that predicts the disease risk reduction.

Other papers study patient adherence to treatment programs based on algorithms. [Lin et al. \(2022\)](#) investigate patients' adherence to an algorithm's advice on type 1 diabetes self-management using a field experiment. [Boutilier et al. \(2022\)](#) demonstrate that personal outreach messages targeting patients with high risk predicted by ML can increase medication adherence for tuberculosis in Kenya.

Lastly, our research is related to the literature that applies predictions and examines their impact on healthcare operations. [Bertsimas and Pauphilet \(2023\)](#) use prediction-based robust optimization to design hospital-wide inpatient flow. [Chan et al. \(2023\)](#) predict the macronutrient content of human milk donations and optimally combine them in pools before distributing them. [Hu et al. \(2025\)](#) propose a prediction-driven staffing framework and evaluate the benefit of surge staffing in the emergency department.

## 2.2. Optimization of Colorectal Cancer Screening

In the management science literature, studies attempt to optimize CRC screenings to maximize the benefits. [Leshno et al. \(2003\)](#) assess different screening policies using a Markov decision process that simulates the evolution of CRC and find that colonoscopy followed by a 10-year interval of follow-up screening, similar to the current guidelines ([U.S. Preventive Services Task Force 2021](#)), achieves the best outcome in terms of reducing mortality. Similarly, [Erenay et al. \(2014\)](#) develop a Markov decision process to optimize colonoscopy screening policies to maximize total quality-adjusted life years. They provide personalized screening policies based on gender and history of CRC and polyp. [Gao et al. \(2025\)](#) study the optimal cutoff value for fecal immunochemical tests that maximizes follow-up colonoscopy adherence, incorporating the screening adherence behavior as an endogenous factor in their model. They show that their optimal cutoff for the fecal immunochemical test could enhance screening effectiveness with fewer follow-up colonoscopies. Our paper sheds light on this stream of literature by investigating the effectiveness of ML-guided CRC screenings on CRC detection, screening adherence, and mortality.

## 2.3. Effects of Colonoscopy Screening

Numerous medical studies investigate the effectiveness of colonoscopy in the prevention and treatment of CRC. Beginning from the National Polyp Study ([Winawer et al. 1993](#)), observational studies find that colonoscopy is associated with a reduction in CRC incidence rates and mortality ([Winawer et al. 1993](#), [Kahi et al. 2009](#), [Brenner et al. 2014](#)). In addition, randomized clinical trials in Europe and the U.S. evaluate the causal effects of colonoscopy.

Although colonoscopy is widely endorsed in the U.S. and Canada, European guidelines currently recommend fecal immunochemical testing as the preferred screening test for CRC because colonoscopy is invasive and expensive and entails a risk of complications ([Bretthauer et al. 2016](#), [European Commission 2022](#)). To evaluate the benefits and harms of colonoscopy, there are ongoing randomized trials in Europe, such as the COLONPREV study in Spain ([Quintero et al. 2012](#), [Castells and Quintero 2015](#)), the NordICC study in Poland, Norway, Sweden, and the Netherlands ([Bretthauer et al. 2016, 2022](#)), and SCREESCO in Sweden ([Forsberg et al. 2022](#)). In the interim reports from these investigations unanimously find that the colonoscopy was effective in detecting adenomas and CRC, achieving higher detection rates than the fecal immunochemical testing ([Quintero et al. 2012](#), [Bretthauer et al. 2016, 2022](#), [Forsberg et al. 2022](#)). In a 10-year follow-up report from the NordICC study, [Bretthauer et al. \(2022\)](#) find that colonoscopy participants had 31% lower CRC incidence rates and 50% lower CRC-related death rates than the usual-care group.

In the U.S., a similar randomized clinical trial is ongoing. Colonoscopy vs. Fecal Immunochemical Test in Reducing Mortality from Colorectal Cancer (CONFIRM) study recruited 50,000 participants in 2017 to test whether colonoscopy is superior to fecal immunochemical test. As the largest

intervention trial conducted by Veterans Affairs, the study will analyze outcomes such as mortality, CRC incidence, and colonoscopy complication rates, following each participant for ten years. The study is expected to be completed in 2027 (Dominitz et al. 2017).

Our study contributes to this stream of literature by analyzing the effects of ML-guided proactive colonoscopy screenings on mortality and providing novel evidence of the effects of colonoscopy on mortality.

#### 2.4. Colonoscopy Patient Outreach and Participation

Despite the evidence showing the effectiveness of colonoscopy screening in detecting CRC (Quintero et al. 2012, Bretthauer et al. 2016, Forsberg et al. 2022), participation in colonoscopy screening remains low. For example, in the NordICC study, only 40% of the eligible colonoscopy group underwent screening (Bretthauer et al. 2016). In the SCREESCO study, 35.1% of people who received an invitation for colonoscopy participated (Forsberg et al. 2022). In the U.S., 46% of eligible adults were overdue for colonoscopy in 2021 (American Cancer Society 2023). Therefore, studying ways to increase adherence to colonoscopy screening is essential.

Previous research finds that patient navigation is effective in increasing colonoscopy participation. Rice et al. (2017) study the effects of a statewide intervention, the New Hampshire Colorectal Cancer Screening Program’s patient navigation, on colonoscopy completion. They find that the patient navigation increased colonoscopy uptake by 11.2 times compared to non-navigated patients and achieved zero missed appointments. This effective intervention was resource-intensive: trained nurse navigators spent an average of 124.3 minutes delivering the patient navigation protocols to each patient (DeGroof et al. 2019). Similarly, other research finds that patient navigation is effective in increasing colonoscopy participation in different settings (Jandorf et al. 2005, Percac-Lima et al. 2009, Honeycutt et al. 2013). Conversely, Leone et al. (2013) find the patient navigator program did not significantly increase CRC screening uptake in 6 months among Medicaid beneficiaries in North Carolina. Besides the patient navigator program, research also investigates the effect of different patient outreach interventions using invitation letters (Gupta et al. 2013, Singal et al. 2017, Basch et al. 2006, Phillips et al. 2015), phone calls (Basch et al. 2006, Phillips et al. 2015).

In line with previous research, our paper investigates the effect of a patient outreach program on colonoscopy screening uptake. Our research is distinct from others because the individuals selected to be contacted by the nurse outreach were determined to have high CRC risk by an ML algorithm. Our nurse outreach informed the patients that they were overdue to get their standard colonoscopy screening and that *additional analyses* had identified them at higher risk than normal (Underberger et al. 2022). It is unclear whether such a program would effectively increase colonoscopy uptake. Moreover, unlike other studies that focus on screening compliance as the outcome, we evaluate the effects of the outreach program on patient outcomes such as mortality and healthcare utilization, in addition to screening uptake.

### 3. Study Setting

Geisinger, headquartered in Danville, PA, is a large healthcare system that operates ten hospital campuses, 133 primary care and specialty clinics, and a health plan with more than half a million members, hiring 25,000 employees with 1,700 physicians. Since 2019, Geisinger has been using a predictive ML algorithm to identify and reach out to patients who may be at higher risk for having or developing CRC. The program assessed every patient within Geisinger’s primary care aged 50-75 years old – the age range [U.S. Preventive Services Task Force \(2016\)](#) recommended for CRC screening – who were not compliant with CRC screening recommendations and had outpatient complete blood count (CBC) results within the previous seven days. The program rolled out weekly by analyzing new input data of patients who had an outpatient CBC test in the previous seven days. The program’s primary intervention was to engage patients who were non-compliant with CRC screening recommendations and who were identified as high-risk according to an ML algorithm by providing direct outreach by a nurse navigator to advise and educate these individuals about the benefits of undergoing a colonoscopy ([Underberger et al. 2022](#)).

The predictive ML algorithm used to identify patients for outreach was developed by a health-tech company, Medial EarlySign. It identifies high-risk patients by analyzing features generated by various moments of age, gender, and CBC indices (e.g., hemoglobin, red blood cells, white blood cells, and platelets) from a blood test taken at an outpatient visit within the last seven days. If available, 3 years of the patient’s past CBC results are also included because historic CBC data improves the prediction accuracy ([Kinar et al. 2016](#), [Underberger et al. 2022](#)). Using these features, a shallow ML model based on XGBoost, a gradient-boosting model, calculates a risk score to determine the likelihood of the presence of CRC. The initial model validation showed that the algorithm sensitivity for detecting CRC within 180 days was 32.9% at 97.5% specificity ([Underberger et al. 2022](#)). [Kinar et al. \(2016, 2017\)](#) provide details of the development of the model. The model was validated at healthcare organizations in the United States ([Hornbrook et al. 2017](#), [Schneider et al. 2020](#)) and the United Kingdom ([Birks et al. 2017](#)).

Among the program’s participating patients, 35% were covered under the Geisinger Health Plan, a health maintenance organization (HMO) where patients must seek care within the Geisinger health system. The remaining patients are either covered under different health plans or are uninsured and may seek care at non-Geisinger facilities, even though their primary care provider (PCP) is within the Geisinger network.

The program flagged only those with the predicted risk score exceeding a fixed threshold for clinical follow-up. Patients were flagged for intervention if their predicted risk score was greater than or equal to 0.150. This cutoff was determined by considering the number of the target population who were overdue for CRC screening and completed outpatient blood tests each week, as well as

the capacity of nurse coordinators and the screening department (Underberger et al. 2022). The sensitivity and specificity at this cutoff are 27.7% and 97.5%, respectively. For patients flagged by the algorithm, a nurse coordinator reviewed the medical chart and removed patients who were actively receiving chemotherapy, had died, or were admitted to the hospital. After the medical chart review, the nurse coordinator contacted the patient by phone. During a phone call, the nurse coordinator provided the following information: i) they were working with the patient’s PCP office, ii) the patient was due for a colonoscopy, iii) an “additional analysis” had identified the patient to be at higher risk than normal for several intestinal conditions (further explained as ulcers, colon polyps, or possibly colon cancer, if the patient asks), and iv) a colonoscopy is the preferred screening method. If the patient agreed, the nurse coordinator scheduled a colonoscopy; otherwise, the patient was offered an appointment with their PCP to discuss other options. Lastly, if the nurse coordinator was unable to reach the patient after three attempts, a certified letter was sent to the patient notifying them of their increased risk for CRC and providing a callback number if they wished to schedule a colonoscopy or talk to their PCP about a colonoscopy (Underberger et al. 2022).

Geisinger allocated significant capacity and resources to run this program. The amount of resources were kept constant during the entire study period. About 450 risk scores (cases) were analyzed each week using complete blood count records from outpatient visits within the last seven days. On average, 11.3 patients were flagged by the ML-algorithm (risk score  $\geq 0.15$ ) for outreach by the nurse coordinator each week. Geisinger allocated 16 hours weekly for a nurse coordinator to perform phone calls and outreach. The gastroenterology department allocated three to five weekly colonoscopy appointments for flagged patients (Underberger et al. 2022).

#### 4. Data

We obtained deidentified data on all participants of Geisinger’s ML-guided CRC screening outreach program between July 2019 and December 2022 and hospital visit records from July 2019 to December 2023. The program’s ML model was not updated during this study period. For each case (i.e., risk score), we observe the patient’s demographics, including birth date, death date (if applicable – death information is directly linked into the Geisinger Electronic Medical Record weekly from the Social Security Administration Death Master File), sex, race, ZIP Code, and insurance type; information related to predicted risk score such as the risk score value, the date when the risk score was calculated, and whether the case (patient-risk score) was flagged or not; information related to outreach by the nurse coordinator including the outreach date and outreach result; data on colonoscopy that include the date of colonoscopy and whether the screening was positive for CRC; and information related to visits to hospitals and clinic offices, including specialty (e.g., oncology, internal medicine), visit date, and admission and discharge dates.

From the original sample of 72,369 cases, we excluded flagged cases that had a colonoscopy before outreach occurred or where the patient was not contacted after a chart review for reasons such as the patient was actively receiving chemotherapy, had died, or was admitted to the hospital. Next, we excluded patients aged 50 because adults who turn 50 are more likely to undergo a colonoscopy, as the [U.S. Preventive Services Task Force \(2016\)](#) recommended screening for CRC in all adults aged 50 to 75 years during our time period of analysis. Lastly, we excluded a few cases for patients with multiple risk scores by limiting one risk score per patient in a month. In [Section 6.5](#), we consider different sample constructions, including patients aged 50 or including patients with multiple risk scores, but restricting to one risk score per patient, and show that our results remain consistent. The resulting sample is 62,485 risk scores of 32,782 patients aged 51 to 75. Among them, 60,964 belong to the control group (risk score below the cutoff), and 1,521 belong to the treated group (risk score greater than or equal to the cutoff). [Online Appendix A](#) provides details for the sample selection process. For outcome variables examining 2-year periods, such as the 2-year mortality and 2-year healthcare utilization, we restrict our sample to risk scores calculated before January 2022 so that we observe outcomes for two years following the risk score.

We analyze four sets of outcomes. First, we analyze outcomes related to colonoscopy. We construct indicators for whether a patient had a colonoscopy within 3 and 6 months. If the patient had a colonoscopy, we also measure time from the date of risk score calculation to the date of colonoscopy (we define this as the time to colonoscopy) and whether the colonoscopy was positive for CRC. For the time to colonoscopy and colonoscopy results, we measure these outcomes only for those who had a colonoscopy. Second, we analyze mortality by creating indicators for death within one year and two years from the date of risk score. Lastly, we measure healthcare utilization for inpatient and clinic office visits by examining the total length of stay (LOS) and the number of office visits in 2 years. [Table 1](#) and [Table 2](#) provide summary statistics of covariates and outcomes, respectively, for the whole sample and by group (treated and control).

## 5. Empirical Strategy

### 5.1. Regression Discontinuity (RD) Design

The impact of ML-guided cancer screening outreach programs is challenging for researchers to identify because the treatment (i.e., flagging for outreach) is not randomized. In our setting, the treatment assignment is a function of the risk of having cancer predicted by the ML. In other words, the treated units are not randomly selected for treatment. Therefore, we cannot simply compare the treated and control groups to estimate the unbiased treatment effect.

To overcome this challenge, we employ a sharp RD design that identifies the causal effect due to the discontinuity in being flagged at the cutoff risk score. As previously described, patients were

**Table 1** Summary Statistics and Covariate Balance

	Main sample mean (Standard deviation)			$p$ -value on $t$ - or $\chi^2$ -test	Falsification Test: Discontinuity at treatment cutoff	
	All	Control	Treated		RD estimate	$p$ -value on RD estimate
	(1)	(2)	(3)	(4)	(5)	(6)
Age	62.482 (7.091)	62.363 (7.075)	67.267 (5.965)	< 0.001	0.118	0.827
Female	0.546 (0.498)	0.546 (0.498)	0.517 (0.500)	0.023	-0.026	0.577
Race				0.024		
White	0.948 (0.221)	0.948 (0.222)	0.957 (0.204)		0.020	0.208
Black	0.028 (0.165)	0.028 (0.166)	0.020 (0.141)		-0.004	0.725
Asian	0.009 (0.092)	0.008 (0.092)	0.013 (0.114)		0.000	0.967
Ethnicity				0.014		
Not Hispanic or Latino	0.955 (0.208)	0.954 (0.209)	0.970 (0.171)		-0.014	0.354
Insurance				< 0.001		
Commercial	0.652 (0.476)	0.655 (0.475)	0.531 (0.499)		-0.023	0.620
Medicare	0.289 (0.453)	0.286 (0.452)	0.415 (0.493)		-0.007	0.868
Medicaid	0.050 (0.218)	0.050 (0.219)	0.045 (0.207)		0.015	0.439
Self-pay	0.001 (0.030)	0.001 (0.029)	0.001 (0.036)		0.001	0.390
Observations	62,485	60,964	1,521			

*Notes.* Columns 1, 2, and 3 show means and standard deviations for all, control group, and treated group, respectively. Column 4 reports the  $p$ -value on the  $t$ -test for continuous variables and chi-squared ( $\chi^2$ ) test for categorical variables. Columns 5 and 6 report estimated RD coefficient and  $p$ -value using the bandwidth of column 2 in [Table 3](#) with a triangular kernel and control variables, except for the corresponding variable. Results are robust to using MSE optimal bandwidths for the corresponding variable ([Calonico et al. 2014](#)).

flagged only if the predicted risk score was greater than or equal to the cutoff point, which we denote by  $c$ . To analyze the impact of ML-guided cancer risk flagging, we estimate the following equation restricting the risk score to be within a bandwidth,  $[c - h_-, c + h_+]$ :

$$y_i = \alpha + \tau FFLAG_i + f(X_i) + \mathbf{Z}'_i \boldsymbol{\beta} + \varepsilon_i, \quad \text{for } c - h_- \leq X_i \leq c + h_+, \quad (1)$$

where  $y_i$  is the outcome of interest (e.g., the likelihood of colonoscopy uptake, mortality) for individual-risk score  $i$ ;  $X_i$  is the running variable, the risk score predicted by the ML,  $FLAG_i$  is an indicator equal to 1 if the predicted risk score is greater than or equal to the cutoff point,  $c$ . That is,  $FLAG_i = \mathbb{1}(X_i \geq c)$ , where  $\mathbb{1}(\cdot)$  is an indicator function;  $f(\cdot)$  is a  $K$ -order polynomial on the running variable and its interaction with the treatment,  $\sum_{k=1}^K (\gamma_{-,k}(X_i - c)^k + \gamma_{+,k} FFLAG_i \times (X_i - c)^k)$ . Because higher-order polynomials tend to produce overfitting and lead to unreliable results near boundary points ([Gelman and Imbens 2019](#), [Cattaneo et al. 2020](#)), we use the local linear ( $K =$

**Table 2** Summary Statistics of Outcome Variables

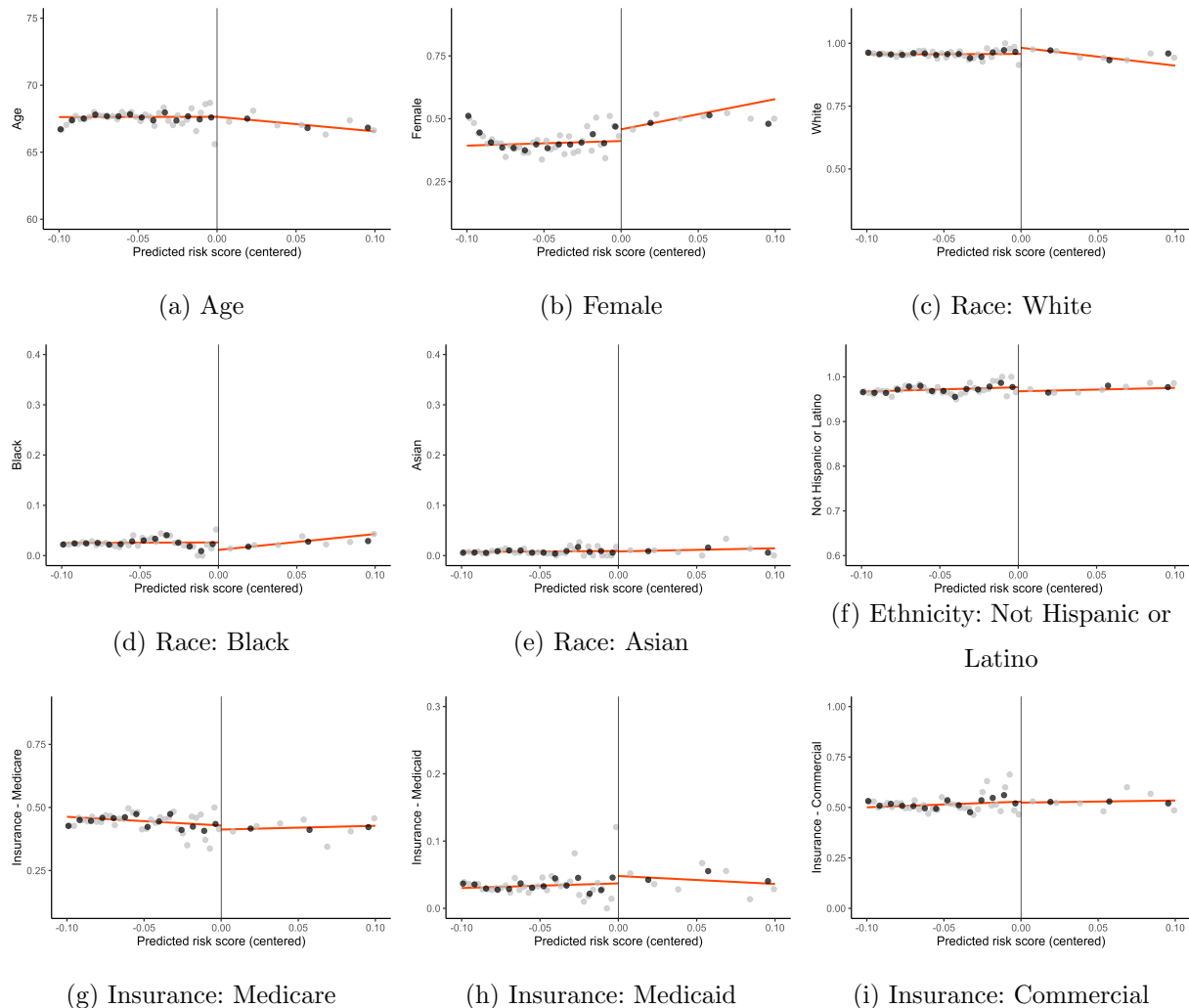
	Sample mean (Standard deviation)			<i>p</i> -value on <i>t</i> -test (4)
	All (1)	Control (2)	Treated (3)	
Colonoscopy in 3 months	0.035 (0.183)	0.033 (0.178)	0.112 (0.315)	< 0.001
Colonoscopy in 6 months	0.062 (0.241)	0.060 (0.237)	0.160 (0.367)	< 0.001
Time to colonoscopy	312.40 (313.90)	318.03 (315.57)	202.32 (255.73)	< 0.001
Positive CRC	0.011 (0.105)	0.008 (0.091)	0.064 (0.246)	< 0.001
1-year mortality	0.046 (0.209)	0.045 (0.207)	0.078 (0.268)	< 0.001
Observations	62,485	60,964	1,521	
2-year mortality	0.080 (0.271)	0.079 (0.269)	0.121 (0.326)	< 0.001
Total LOS in 2 years	0.597 (3.482)	0.589 (3.451)	0.900 (4.529)	0.019
Number of office visits in 2 years	10.569 (10.127)	10.503 (10.060)	13.140 (12.185)	< 0.001
Observations	46,994	45,812	1,182	

*Notes.* For 2-year period outcomes, risk scores calculated before January 2022 are included.

1) and quadratic ( $K = 2$ ) polynomials, following [Gelman and Imbens \(2019\)](#)’s recommendation. We use MSE optimal bandwidths separately for each side of the cutoff,  $h_-$  and  $h_+$ , because the conditional variance of outcomes given the score near the cutoffs are different for each side, and the curvature of the unknown regression functions could be different for each side ([Calonico et al. 2014](#), [Cattaneo et al. 2020](#)). We additionally use a triangular weighting kernel as it leads to an MSE-optimal point estimate ([Cattaneo et al. 2020](#), p. 37). In [Section 6.5](#), we test the robustness of our results by changing the bandwidths and using different kernel functions. A vector of covariates,  $\mathbf{Z}_i$ , includes age and indicators for females, race, and insurance types. Here,  $\tau$  is the RD estimator that captures the average outcome jump at the cutoff in a fully interacted local polynomial regression fit after controlling the effect of the covariates.

## 5.2. Validity of RD Design

The identifying assumption of the treatment effect in [Equation \(1\)](#) is that the regression functions  $\mathbb{E}[Y_i(0)|X_i = x]$  and  $\mathbb{E}[Y_i(1)|X_i = x]$  are continuous in  $x$  at  $c$  ([Hahn et al. 2001](#)), where  $Y_i(0)$  and  $Y_i(1)$  denote the outcome under control and treatment, respectively. Here, we are adopting the continuity-based framework instead of the local randomization framework (See [Cattaneo and Titiunik \(2022\)](#) and [Sekhon and Titiunik \(2017\)](#) for the distinction between the two frameworks). This implies that other factors of the outcome are continuous with respect to  $X_i$  ([Lee and Lemieux](#)

**Figure 2** Balance of Predetermined Characteristics

*Notes.* The figure plots the characteristics of the sample. Each figure depicts data points within 0.10 risk scores around the cutoff. Black points present the average value of the specified variable for risk scores with 20 bins on each side of the cutoff. Gray points present the average value of the specified variable for risk scores with 50 bins on each side of the cutoff. The solid, red trend lines are the predicted values from local linear regression fits of the specified variable on the score, fitted separately above and below the cutoff, that uses a triangular kernel and a bandwidth of 0.10 on each side. The solid vertical line represents the cutoff risk score value.

2010). We test this assumption by examining whether treated and control units near the cutoff are similar in predetermined characteristics. We run this falsification test the same way as the outcome of interest, replacing  $y_i$  in Equation (1) with the observed characteristics. In other words, we test whether the treatment affects predetermined covariates. Figure 2 presents a graphical analysis with linear regression lines. Online Appendix B Figure OA.1 presents a graphical analysis using quadratic regressions. The characteristics, including age, sex, race, ethnicity, and insurance types, do not show apparent significant discontinuities at the treatment cutoff. Table 1 columns 5 and 6 present the test results that fail to reject that there is no discontinuity in each of the covariates at the cutoff.

Another identification requirement is no manipulation of the forcing variable (i.e., risk score) around the cutoff point (Lee and Lemieux 2010). This is a plausible assumption in our setting because the risk score is predicted solely by a pre-trained ML model; therefore, no individual can manipulate the risk score. As Figure OA.3(a) shows, there is no bump in distribution around the cutoff. Still, we formally test this assumption, and the McCrary (2008) test for breaks in the density of the running variable fails to reject the null hypothesis of no discontinuity in the density at the cutoff (Figure OA.3(b)).

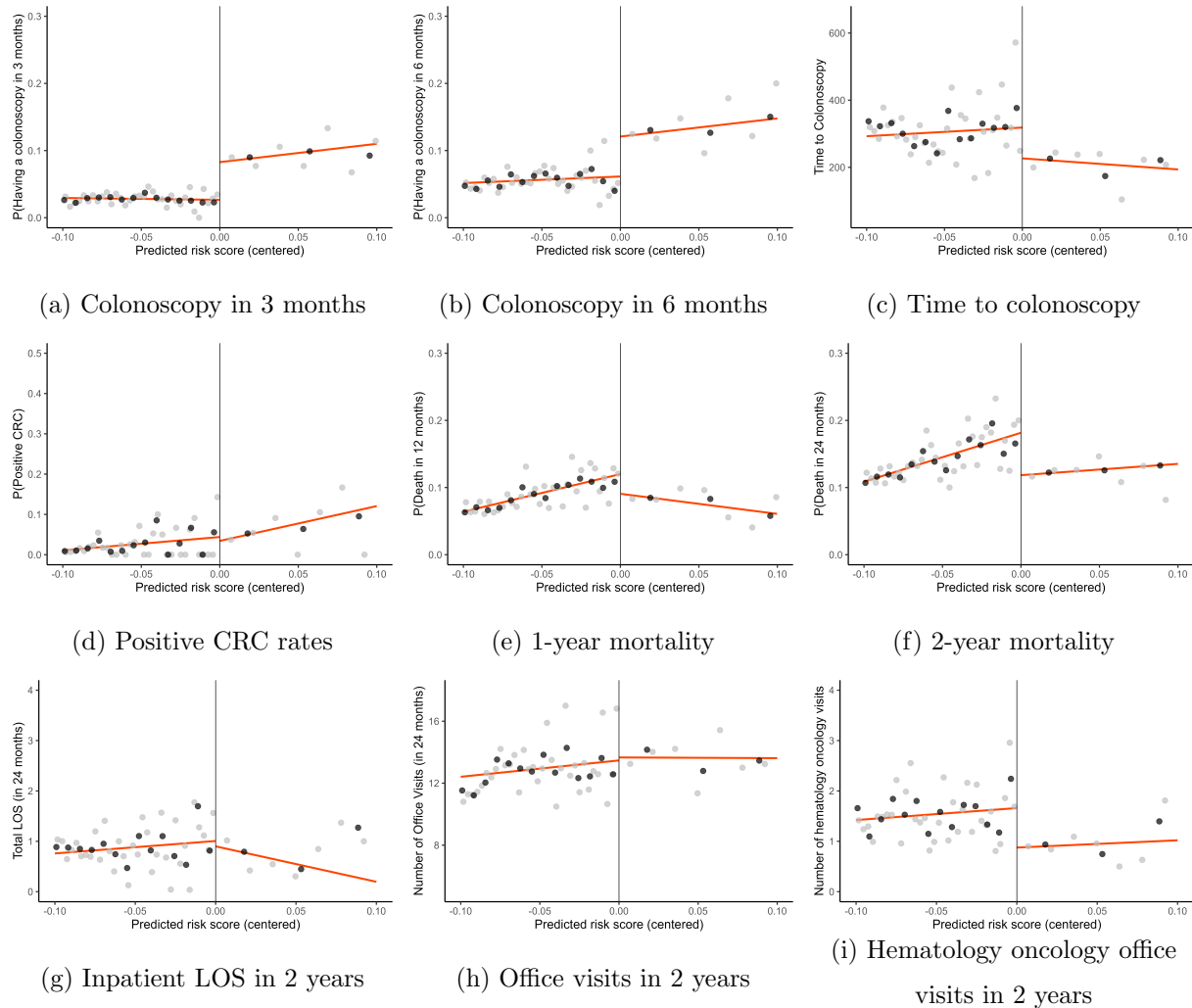
## 6. Results

In this section, we present the results of the analysis of the effects of the ML-guided CRC screening outreach program on colonoscopy uptake, mortality, and healthcare utilization using the RD design. We formally estimate the treatment effects by estimating Equation (1), and we present RD plots with linear regressions in Figure 3 as they are informative about how the treatment effect is identified around the cutoff point (Lee and Lemieux 2010). For the RD plots using quadratic regressions, we display them in Online Appendix B. In our estimation results, we present them using both linear (odd numbered columns) and quadratic regressions (even numbered columns).

### 6.1. Colonoscopy

We first examine the change in the likelihood of having a colonoscopy screening at the cutoff (Table 3). We analyze two outcomes, whether a patient had a colonoscopy in 3 months (columns 1 and 2) and 6 months (columns 3 and 4) from the date the risk score was calculated. Because our sample consists of patients due for colonoscopy based on the national guidelines, those in the control group (not flagged) could also get a colonoscopy. Column 1 indicates that the program significantly increases the likelihood of getting a colonoscopy in 3 months by about 6.0 pp ( $p = 0.001$ ) at the cutoff, a 214% increase relative to the control group within the bandwidth. Results in column 3 also show positive effects at the cutoff, about 6.9 pp ( $p = 0.002$ , 117% increase), on the chance of having a colonoscopy in 6 months. We also analyze the time to colonoscopy, conditional on having a colonoscopy (columns 5 and 6). We find that the program significantly decreases the time to have a colonoscopy by about 124 days ( $p = 0.067$ , 39% decrease) at the cutoff. This is a conservative estimate because we do not include patients who did not have colonoscopy during the study period and the control group has significantly lower colonoscopy uptake compared to the treated group. Overall, the results suggest that the program significantly changed the behavior of the flagged patients to have colonoscopies earlier.

Figure 3 RD Plots for Outcomes



*Notes.* Each figure depicts data points within 0.10 risk scores around the cutoff. Black points present the average value of the specified variable for risk scores with 20 bins on each side of the cutoff. Gray points present the average value of the specified variable for risk scores with 50 bins on each side of the cutoff. The solid, red trend lines are the predicted values from local linear regression fits of the specified variable on the score, fitted separately above and below the cutoff, that uses a triangular kernel and a bandwidth of 0.10 on each side. The solid vertical line represents the cutoff risk score value. CRC: colorectal cancer, LOS: length of stay.

## 6.2. Screening results

Next, we analyze the results of colonoscopy screenings. Figure 1(a) plots the positive diagnosis rates for CRC by group for the whole sample. For the treated group (flagged cases), the average positive rate is 6.42%, significantly larger than the control group's positive rate of 0.84% ( $t$ -test  $p$ -value  $< 0.001$ ). This is consistent with Underberger et al. (2022), who evaluated the diagnosis rates of flagged patients for the same program at different periods. The positive rate for the control group, 0.84% (95% CI: [0.63%, 1.06%]), is consistent with 1.0% positive rates from asymptomatic colonoscopy screenings (Lieberman et al. 2000). Although the significantly larger average positive

**Table 3 Effects on Colonoscopy at the Cutoff**

Outcome	$P(\text{Colonoscopy in 3 months})$		$P(\text{Colonoscopy in 6 months})$		Time to Colonoscopy	
	Linear (1)	Quadratic (2)	Linear (3)	Quadratic (4)	Linear (5)	Quadratic (6)
Flag	0.060** (0.018)	0.059** (0.021)	0.069** (0.023)	0.074** (0.026)	-124.00 <sup>+</sup> (67.64)	-189.56 <sup>+</sup> (114.71)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.070, 0.073]	[-0.088, 0.112]	[-0.070, 0.077]	[-0.108, 0.111]	[-0.053, 0.068]	[-0.048, 0.083]
Mean below cutoff	0.028	0.029	0.059	0.052	315.12	327.45
Observations	6,268	11,953	6,371	24,539	433	413
Control	5,460	10,971	5,550	23,557	276	235
Treated	808	982	821	982	157	178

*Notes.* The table presents the estimates using local linear regression (columns 1, 3, and 5) or local quadratic regression (columns 2, 4, and 6) with a triangular weighting kernel. Control variables include age, sex, race, and insurance type (Medicare, Medicaid, military, commercial, self-pay). Bandwidth shows the interval centered around the cutoff point using MSE optimal bandwidths (Calonico et al. 2014). The observations row shows effective observations within the bandwidth. Robust standard errors are reported in parentheses. Total observations: 62,485 (columns 1 to 4) include our main sample; 7,366 (columns 5 to 6) include risk scores that underwent a colonoscopy. <sup>+</sup> $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

rates of the flagged group demonstrate the predictive power of the ML model, there should not be a discontinuity in the positive rates at the cutoff because the continuous risk score indicates a smoothly changing risk of CRC. In other words, the program should not affect the chance of having CRC. Figure 3(d) reflects this aspect, illustrating smoothly increasing positive CRC rates in predicted risk score but no discontinuity at the cutoff. Table 4, columns 1 and 2 confirm that the treatment effects at the cutoff are not significantly different from 0 ( $p = 0.529$  in column 1;  $p = 0.538$  in column 2).

**Table 4 Effects on CRC and Mortality at the Cutoff**

Outcome	Positive CRC		1-year mortality		2-year mortality	
	Linear (1)	Quadratic (2)	Linear (3)	Quadratic (4)	Linear (5)	Quadratic (6)
Flag	-0.043 (0.068)	-0.059 (0.096)	-0.038 <sup>+</sup> (0.020)	-0.039 (0.024)	-0.062* (0.027)	-0.072* (0.033)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.032, 0.043]	[-0.042, 0.083]	[-0.094, 0.073]	[-0.121, 0.098]	[-0.077, 0.096]	[-0.100, 0.145]
Mean below cutoff	0.034	0.039	0.081	0.064	0.143	0.128
Observations	244	356	14,723	37,055	6,320	14,191
Control	118	179	13,915	36,127	5,614	13,345
Treated	126	177	808	928	706	846

*Notes.* The table presents the estimates using local linear regression (columns 1, 3, and 5) or local quadratic regression (columns 2, 4, and 6) with a triangular weighting kernel. Control variables include age, sex, race, and insurance type (Medicare, Medicaid, military, commercial, self-pay). Bandwidth shows the interval centered around the cutoff point using MSE optimal bandwidths (Calonico et al. 2014). The observations row shows effective observations within the bandwidth. Robust standard errors are reported in parentheses. Total observations: 7,366 (columns 1 and 2) include risk scores that underwent a colonoscopy; 62,485 (columns 3 to 4) include our main sample; 46,994 (columns 5 and 6) include risk scores collected until December 2021. <sup>+</sup> $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

### 6.3. Mortality

We study the effects on mortality by looking at whether the patient died in one year and two years following the date the risk score was calculated. [Table 4](#), columns 3 to 6 present the results. For the follow-up period of 1 year, we find that the RD estimate using linear regression is negative and marginally significant in column 1 (estimate = -0.038,  $p = 0.050$ ) but becomes not statistically significant when using quadratic regression in column 2 (estimate = -0.039,  $p = 0.106$ ). The effect on mortality becomes statistically significant at two years. The program significantly reduces 2-year mortality at the cutoff by about 6.2 pp ( $p = 0.022$ , 43% decrease) estimated using linear regression and by about 7.2 pp ( $p = 0.031$ , 56% decrease) estimated by quadratic regression. This finding is the opposite of the simple comparison of mortality between the whole treated and the whole control group, i.e. not restricting the sample to those that fall within the MSE-optimal bandwidth. On average, the treated group (12.1%) has higher 2-year mortality than the control group (7.90%,  $t$ -test  $p$ -value  $< 0.001$ ), as [Figure 1\(b\)](#) shows. This is because the treated group is older and has a higher risk of CRC, as predicted by the ML. On the other hand, the estimated local treatment effect at the cutoff indicates a reversed relationship in which treated units near the cutoff have significantly lower mortality than the control units near the cutoff. In [Section 6.5](#), we examine the robustness of our results. We discuss potential explanations of the effects on mortality in [Section 7](#).

### 6.4. Healthcare Utilization

We analyze the effects on healthcare utilization as a last set of outcome variables. We analyze two types of utilization: inpatient and clinic office visits. There is a potential censoring issue for healthcare utilization because we observe utilization only if the patient uses services within Geisinger. Our data does not include utilization of services outside our partner organization. This potential censoring of utilization does not exist for patients in the Geisinger Health Plan HMO. As such, for the healthcare utilization outcomes, we conduct a subset analysis for just Geisinger Health Plan patients for whom we observe all utilization in [Table 5](#). First, we analyze 2-year LOS for any type of inpatient hospitalization. In columns 1 and 2, the estimates are not statistically significantly different from 0 (estimate = -0.063,  $p = 0.944$  in column 1; estimate = 0.014,  $p = 0.990$  in column 2), suggesting that there is no significant gap in 2-year LOS at the cutoff. Next, we analyze the number of all-type clinic office visits, except for gastroenterology, because colonoscopies fall under this specialty (columns 3 and 4). We do not find that the program significantly impacts the total number of office visits in two years at the cutoff (estimate = 0.752,  $p = 0.646$  in column 3; estimate = -0.426,  $p = 0.832$  in column 4).

We further analyze inpatient LOS and office visits by type of specialty (e.g., intensive care, radiology, etc.), and we do not find significant discontinuities at the cutoff, except for one specialty:

**Table 5** Effects on Healthcare Utilization (2 years) at the Cutoff

Outcome	Inpatient LOS		Number of office visits		Hematology oncology office visits	
	Linear (1)	Quadratic (2)	Linear (3)	Quadratic (4)	Linear (5)	Quadratic (6)
Flag	-0.063 (0.904)	0.014 (1.110)	0.752 (1.637)	-0.426 (2.004)	-1.006 <sup>+</sup> (0.521)	-1.076 <sup>+</sup> (0.586)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.057, 0.074]	[-0.079, 0.112]	[-0.056, 0.077]	[-0.052, 0.137]	[-0.059, 0.046]	[-0.083, 0.112]
Mean below cutoff	0.892	0.863	13.153	13.116	1.481	1.574
Observations	1,091	2,152	1,047	1,038	1,103	2,469
Control	896	1,908	849	764	952	2,226
Treated	195	244	198	274	151	243

*Notes.* The table presents the estimates using local linear regression (columns 1, 3, and 5) or local quadratic regression (columns 2, 4, and 6) with a triangular weighting kernel. Control variables include age, sex, and race. Bandwidth shows the interval centered around the cutoff point using MSE optimal bandwidths (Calonico et al. 2014). The observations row shows effective observations within the bandwidth. Robust standard errors are reported in parentheses. Total observations: 15,633 include Geisinger Health Plan holders with risk scores collected until December 2021. <sup>+</sup> $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

hematology oncology office visits. For hematology oncology office visits, we find a marginally significant decrease in the number of office visits in 2 years by 1.006 when using linear regression in column 5 ( $p = 0.053$ , 68% decrease) and by 1.076 when using quadratic regression in column 6 ( $p = 0.066$ , 68% decrease) at the cutoff. This finding may suggest that the program could have spillover effects on diagnosing other cancers early, as the research validating the ML model finds (Kinar et al. 2016), which could potentially contribute to a reduction in overall mortality. However, we cannot fully validate this as we do not have data on the stage of detected cancers other than CRC. For brevity, we present the LOS and office visit results of specialties other than hematology oncology in Table OA.2 and Table OA.3, respectively.

## 6.5. Robustness Checks

**6.5.1. Parametric RD** So far, we report results using a local polynomial approach by estimating regression lines using observations within an MSE optimal bandwidth around the cutoff. Because the number of observations within the bandwidth to the right of the cutoff is relatively low in our data, the estimated regression line on the right side of the cutoff could be sensitive to noise. To address this concern, we use a parametric RD by imposing an identical second-order regression line in risk scores on both sides. That is, we set  $f(X_i)$  in Equation (1) as  $\sum_{k=1}^2 \gamma_k (X_i - c)^k$ , without the bandwidth around the cutoff. We use a second-order polynomial, rather than the first-order, based on Akaike’s criterion (Lee and Lemieux 2010). We present results in Table 6 using all risk scores in the entire range, the top 25% risk scores, and the top 10% risk scores to examine the sensitivity of the results to different risk score ranges.

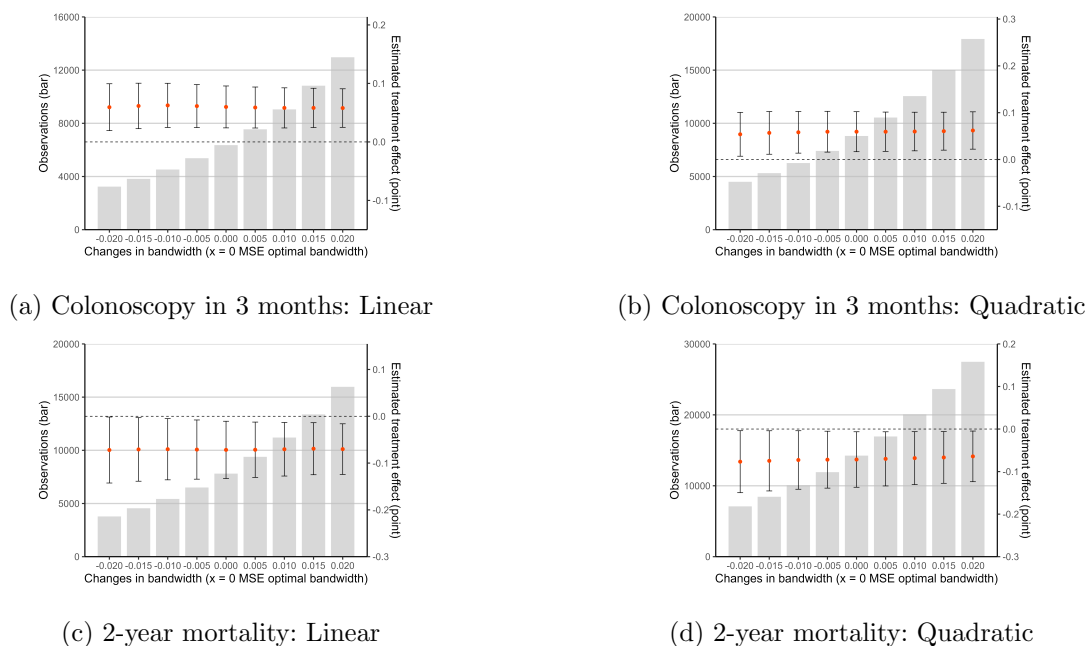
We find consistent findings with the results from the nonparametric RD. For colonoscopy uptake in 3 months and 6 months, we find statistically significant positive jumps at the cutoff, consistently across different ranges of risk scores. The magnitudes of the coefficients are similar to the estimates from the nonparametric RD in Table 3. For the likelihood of having a colonoscopy in 3 months, the RD estimates range between 5.9 pp to 6.5 pp increase at the cutoff, depending on the included sample risk scores. Likewise, for the likelihood of having a colonoscopy in 6 months, the estimates range between 5.8 pp to 7.4 pp increase at the cutoff, depending on the included sample risk scores. Next, for 2-year mortality, we find statistically significant negative drops at the cutoff. The magnitude of the RD estimates decreases as we restrict the sample to higher risk scores, from -11.8 pp for all risk scores to -4.9 pp for top 10% risk scores. This is because mortality increases as risk scores increase (note the mean values below the cutoff increase as sample risk scores get higher), which affects the regression lines. Overall, our results remain consistent when using parametric RD with different samples.

**6.5.2. Sensitivity Test to Bandwidths** We conduct a sensitivity test of our results to varying bandwidths. Following Cattaneo et al. (2020), we investigate the sensitivity to bandwidths over small ranges around the MSE-optimal bandwidth. Increased bandwidths will increase bias due to the non-random treatment assignment; reduced bandwidths will increase variance due to reduced sample size. The MSE-optimal bandwidth optimizes this tradeoff; deviating too far from the MSE-optimal bandwidth could lead to unreliable point estimates and conclusions. Figure 4 shows the results for the likelihood of having a colonoscopy in 3 months (panel a and b) and 2-year mortality (panels c and d). The RD estimates using different bandwidths (from  $\pm 0.005$  to  $\pm 0.02$ ) with different numbers of observations around the MSE-optimal bandwidth are stable and consistently different from zero.

**Table 6** Effects on Colonoscopy Uptake and Mortality at the Cutoff Using Parametric RD

Outcome	Colonoscopy in 3 months			Colonoscopy in 6 months			2-year mortality		
	(1) All	(2) Top 25% risk scores	(3) Top 10% risk scores	(4) All	(5) Top 25% risk scores	(6) Top 10% risk scores	(7) All	(8) Top 25% risk scores	(9) Top 10% risk scores
Flag	0.064*** (0.010)	0.059*** (0.012)	0.065*** (0.015)	0.074*** (0.012)	0.058*** (0.014)	0.064*** (0.018)	-0.118*** (0.015)	-0.071*** (0.019)	-0.049* (0.021)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Mean below cutoff	0.033	0.028	0.029	0.060	0.053	0.059	0.079	0.131	0.151
Observations	62,485	15,950	6,555	62,485	15,950	6,555	46,994	11,992	4,927

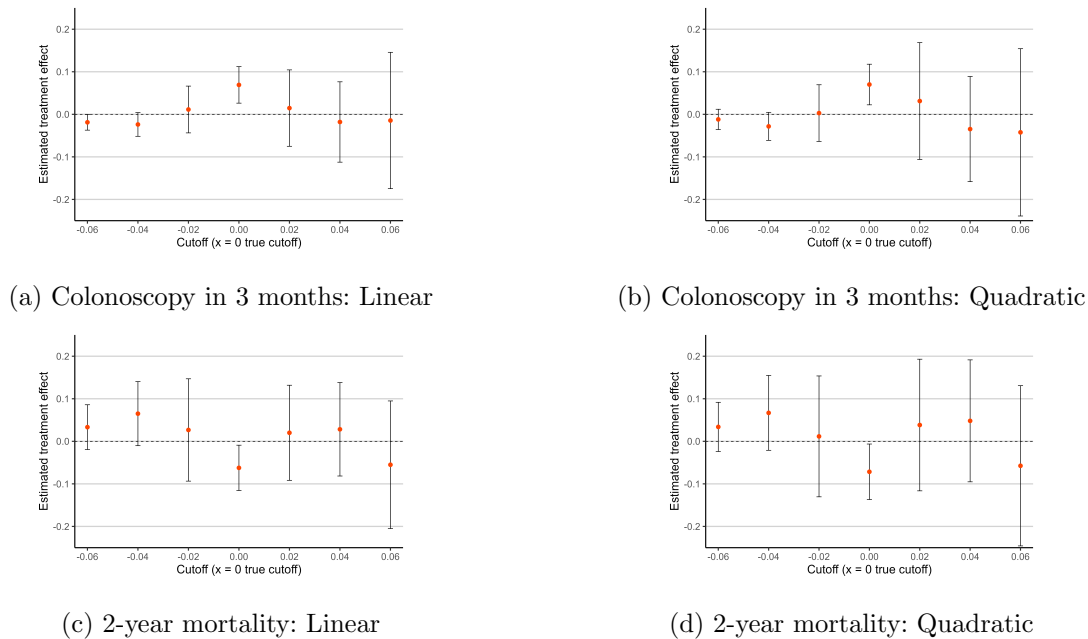
*Notes.* The table presents the estimate using a parametric RD. Control variables include age, sex, race, and insurance type (Medicare, Medicaid, military, commercial, self-pay). Robust standard errors are reported in parentheses. For columns 7 to 9, risk scores collected until December 2021 are included. <sup>+</sup> $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

**Figure 4** Sensitivity to Varying Bandwidth

*Notes.* Figure shows the estimated RD coefficient (point) with a 95% confidence interval (error bar) by changing the bandwidth around the MSE-optimal bandwidth ( $x = 0$ ). We use local linear regression in panels (a) and (c), quadratic regression in panels (b) and (d), and a triangular kernel with control variables for estimation. Bars plot the number of observations within the bandwidth. The dashed horizontal line is the zero line.

**6.5.3. Placebo Cutoffs and Outcomes** Next, as a falsification test, we examine treatment effects at placebo cutoff values to test whether our results reflect fluctuations in outcomes across the predicted CRC risks. If discontinuities in outcomes across different CRC risk scores are common, our estimated treatment effect might be due to the regular fluctuation rather than the effect driven by the program. We assign placebo cutoff points away from the true cutoff and estimate the placebo effect using the control and treatment groups separately. In other words, for placebo cutoffs below the true cutoff, we use only control group observations, and for placebo cutoffs above the true cutoff, we use only treated group observations. This avoids contamination due to real treatment effects (Cattaneo et al. 2020, p. 89). Figure 5 plots the test results. For both outcomes, having a colonoscopy in 3 months (panels a and b) and 2-year mortality (panels c and d), coefficients from placebo tests are not significantly different 0, and only the coefficient at the true cutoff ( $x = 0$ ) is statistically significantly different from 0. The results confirm that our results are not mere reflections of variations in outcomes.

Additionally, we analyze two placebo outcomes not related to CRC – orthopedics and psychiatry 2-year LOS – as another falsification test. The intervention targeting CRC should not significantly affect these unrelated placebo outcomes. Table OA.4 presents RD estimates indicating no significant discontinuities at the cutoff risk score, neither for orthopedics nor psychiatry LOS, which supports that spurious correlations do not drive our results.

**Figure 5** Tests Using Placebo Cutoff Values

*Notes.* Figure shows the estimated RD coefficient (point) with a 95% confidence interval (error bar), using placebo cutoff values around the true cutoff at 0. The estimate at  $x = 0$  is the true estimate at the true cutoff. We use local linear regression in panels (a) and (c), quadratic regression in panels (b) and (d), and a triangular kernel with control variables for estimation. The dashed horizontal line is the zero line.

**6.5.4. Other Kernel Functions** We test the robustness of our results to estimation using different kernel functions other than the triangular kernel: uniform kernel,  $K(\frac{X_i - c}{h}) = \mathbb{1}(|\frac{X_i - c}{h}| \leq 1)$ , and Epanechnikov kernel,  $K(\frac{X_i - c}{h}) = (1 - (\frac{X_i - c}{h})^2)\mathbb{1}(|\frac{X_i - c}{h}| \leq 1)$ . The uniform kernel gives equal weight to all observations whose scores are within the bandwidth, and the Epanechnikov kernel gives a quadratic decaying weight to observations within the bandwidth. We find that the results are not sensitive to the particular choice of the kernel (Table OA.5). The estimates remain consistent in direction, magnitude, and statistical significance.

**6.5.5. Different Samples** Our main sample excludes risk scores of patients aged 50 because the national screening guidelines recommend colonoscopy for all adults starting at age 50. We rerun analyses using all available samples (risk scores), including risk scores of patients aged 50. Table OA.6 presents the results of the analysis of the extended sample. In addition, our main sample allows patients to have multiple risk scores, at most one per month. This could raise a concern about potential bias in RD point estimates due to repeated observations per patient. We rerun the analysis by allowing one risk score (the maximum risk score) per patient to test our results' robustness. The results are consistent with our main sample when we use different samples. At the cutoff, we find significant increases in the likelihood of having a colonoscopy in 3 and 6 months and significant decreases in 2-year mortality.

**6.5.6. Log-transformed Risk Scores** We test the robustness using log-transformed risk scores, as the predicted risks are highly skewed and dispersed on the right side of the cutoff (skewness = 6.25). The log transformation reduces the skewness (skewness = 0.08), and the results remain consistent when using the log-transformed risk scores (Table OA.8).

**6.5.7. Clustered Standard Errors** As our main sample allows the same patient to have multiple risk scores (at different times), the errors could be serially correlated within a patient. To address this concern, we employ cluster-robust variance estimators. This method yields not only different estimated standard errors relative to the unclustered case but also different point estimates because this changes the MSE-optimal bandwidth (Cattaneo et al. 2020, p. 74). Table OA.9 presents the results using clustered standard errors, and the findings remain robust, indicating that the results are valid under potential serial correlation. While clustering accounts for serial correlation in variance estimation, it does not mitigate potential bias in RD point estimates arising from repeated observations per patient. To address this concern, Section 6.5.5 presents an analysis that restricts the sample to one risk score per patient.

**6.5.8. COVID-19 Pandemic** Lastly, as our data include the COVID-19 pandemic period (the program was paused from April to June 2020 due to the pandemic) during which people had high mortality, one may be concerned that the pandemic drives our results on mortality. If this is true, there should be a discontinuity in the chance of having COVID-19 infection at the cutoff. Although this is very unlikely – as supporting evidence, we don’t find significant discontinuity in visits to family medicine offices at the cutoff (Table OA.3) – we address this concern with two approaches: first, we include year dummy variables as control variables in our estimation to control for year effects. Our results remain consistent after controlling the year effects (Table OA.10). Second, we rerun our analysis excluding deaths in the year 2020 (the year with high mortality due to COVID-19) and find consistent results in colonoscopy uptake and mortality (Table OA.11). Overall, this suggests that our results are not driven by the COVID-19 pandemic.

## 7. Discussion and Conclusions

We study the effects of the ML-guided CRC screening outreach program. Using an RD design, we find that the program increased patients’ participation in colonoscopy by 6.0 pp in 3 months (214% increase relative to the control group within the bandwidth) and by 6.9 pp in 6 months (117% increase) and significantly reduced the time to have a colonoscopy by about 124 days (39% decrease) both at the cutoff risk score. Considering that about 46% of U.S. adults aged 45 years and older were overdue for colonoscopy in 2021 (American Cancer Society 2023), this jump in colonoscopy participation is substantial.

Importantly, we find that the program decreased 2-year mortality by 6.2 pp (43% reduction) at the cutoff despite no discontinuity in the risk of having CRC at the cutoff. Given this, if it were possible to roll out this program across the entire United States, the potential number of lives saved would be 263,000. Of course, finding the resources to do this may be prohibitive at such a large scale. For an average-sized PCP practice with 2,500 patients, this could translate to 4.7 lives saved within 2 years. We do not find significant changes in healthcare utilization at the cutoff, including the all-type inpatient LOS and clinic office visits.

*Earlier Colonoscopy* is a potential driver for our results on reduced mortality. As our analysis indicates, the patients just below the cutoff and just above the cutoff have a similar risk of CRC, as the risk score is smooth and continuous around the cutoff (see [Figure 3\(d\)](#)). However, the treated patients around the cutoff are more likely to undergo colonoscopies by 6.0 pp in 3 months and 6.9 pp in 6 months, and significantly earlier (124 days) than those right below the cutoff. Other medical research suggests that earlier screenings will enable earlier diagnosis and treatment, which can lead to higher survival rates. [Tørring et al. \(2013\)](#) find that longer diagnostic intervals were associated with increased mortality for CRC among the Danish patients. Similarly, [Lee et al. \(2019\)](#) conduct a retrospective study of 39,000 CRC-diagnosed patients in Taiwan and find that shorter diagnosis-to-treatment intervals ( $\leq 30$  days) have significantly higher survival rates compared to longer time intervals (31–150 days and  $\geq 151$  days), across all cancer stages. These findings underscore the importance of early screening followed by prompt treatment for improved survival. In addition, precancerous tumors like adenomas or polyps could be removed during a colonoscopy, which is shown to significantly reduce the incidence rates of CRC and mortality ([Winawer et al. 1993](#), [Zauber et al. 2012](#)). Due to the limited number of diagnosed cases, we are unable to rigorously evaluate the program’s effect on time to CRC diagnosis as a mechanism for reduced mortality. Nonetheless, we provide an exploratory analysis of time to CRC diagnosis and CRC stages in [Online Appendix L](#).

*Potential Spillover Effects* of the CRC program may improve outcomes for other cancers because the ML model has some sensitivity (i.e., ability to detect the disease) to other cancers, including stomach cancer, hematological cancers, and to lung cancer even though the current program is designed only for predicting and flagging the risk of CRC ([Kinar et al. 2016](#)). This sensitivity to other cancers is due to their age-dependent incidence, as age is a predictor of the model ([Kinar et al. 2016](#)). Therefore, the risk scores from the model could reflect the risks of cancers, not only for CRC. In the analysis of the hematology-oncology clinic office visits, we find a decrease in the number of visits in two years (estimate = -1.006,  $p = 0.053$ ) at the cutoff. Note that if the intervention detected *more* cancers of other types, this could potentially lead to more office visits. On the other hand, if the cancers would be detected eventually and the intervention resulted in *earlier* detection, this could result in improved prognosis, fewer office visits, and reduce overall mortality.

## 7.1. Contributions to Literature

As discussed in Section 2, existing literature has shown that ML algorithms can predict the risk of having cancers such as CRC (Kinar et al. 2016, 2017, Birks et al. 2017, Hornbrook et al. 2017, Schneider et al. 2020). A separate stream of literature provides some evidence that colonoscopy screening is associated with a reduction in CRC incidence and mortality (Winawer et al. 1993, Kahi et al. 2009, Brenner et al. 2014, Bretthauer et al. 2022). Another branch of research finds that proactive nurse outreach and patient navigation programs could increase colonoscopy participation (Jandorf et al. 2005, Percac-Lima et al. 2009, Honeycutt et al. 2013, Rice et al. 2017).

Our work synthesizes and contributes to these streams of literature by rigorously evaluating a nurse outreach program that was designed to leverage these different insights and intervene on patients who were identified as high risk for CRC by an ML algorithm. We show that a nurse outreach program targeting high-risk patients identified by ML leads to increased colonoscopy uptake, high CRC detection rates, and reduced mortality. Prior work has tackled each of the components (prediction, screening, and proactive outreach) separately. In contrast, our paper demonstrates that a program that integrates ML, screening, and patient outreach could be impactful.

## 7.2. Managerial Implications

**7.2.1. Estimating Unbiased Effects of ML-guided Screening Outreach** Our paper informs managers how to evaluate a similar program/policy by demonstrating an approach to estimating an unbiased causal estimate using the RD design. As most disease prediction models generate a risk score for having a disease and identify high-risk patients based on a fixed threshold score, the RD design used in this paper could be helpful in many settings. Estimating the causal effects of ML-guided programs could be challenging for hospital managers or policymakers because the treatment is not randomized, and the treated group has a high risk of having the disease as classified by the ML. Thus, one cannot (and should not) naively compare outcomes between the groups. For example, if we compare the mortality rate or LOS between the treated group and the control group, as in Table 2, we may wrongly conclude that the program was unsuccessful as the treated group has higher mortality and LOS than the control group, on average.

**7.2.2. Planning Service Capacity** Estimating unbiased causal effects on colonoscopy uptake is also crucial for service capacity planning. A simple comparison between the treated and the control group will overestimate the increase in colonoscopy uptake. For example, by naively comparing the colonoscopy participation in 3 months, one may estimate the effect of 7.9 pp increase (see Table 2), which overestimates the true, *causal* treatment effect from the RD (6.0 pp) by 1.9 pp. Planning colonoscopy capacity for ML-guided screening outreach program based on this overestimate will lead to 19 unused colonoscopy screenings annually per 1,000 participants in the

outreach program. This translates to unused capacity costs of approximately \$40,375 per 1,000 participants per year. (The average cost of colonoscopy in the U.S. was \$2,125 between 2014 and 2019 (Fisher et al. 2022)). Considering the average of 450 patients per week in the program, this implies annual savings of approximately \$944,775 for our partner hospital when capacity planning is based on the unbiased effect rather than the overestimate.

In the initial rollout of this program, our partner hospital, Geisinger, allocated a weekly screening capacity of 3 to 5 colonoscopy appointments for flagged patients (Underberger et al. 2022). Our estimated treatment effect suggests that there would be 1.3 additional colonoscopies each week due to the program’s implementation. Prior to our study, it was not a priori clear how much additional colonoscopy capacity would be needed to support the program. Because the initial rollout provided ample capacity, this allowed us to evaluate the program’s effects on screening participation without censoring colonoscopy demand due to capacity limitations. Our work suggests that 2 ( $> 1.3$ ) additional weekly colonoscopies would be sufficient to support the additional demand introduced by the flagged patients. Thus, Geisinger could potentially achieve the same throughput and life savings with less screening capacity; alternatively, they could serve more flagged patients with the current capacity by lowering the threshold risk score for the flag.

**7.2.3. Improving Colonoscopy Compliance** Although colonoscopy is effective for CRC detection and reducing mortality, compliance has been low because it is invasive, expensive, and resource-intensive (Winawer et al. 1993, Zauber et al. 2012, Shaukat and Levin 2022). Our results demonstrate that an ML-guided screening program could considerably increase colonoscopy uptake by approximately 6%, targeting those overdue for screening and at high risk for CRC. This increase would translate to 3,907,608 more colonoscopies per year nationally. A back-of-the-envelope calculation based on the U.S. Preventive Services Task Force (2021) model estimates that increasing 6% colonoscopy screenings among adults aged 45 years would lead to 20.2 more life-years gained per 1000 individuals based on the estimates from the U.S. Preventive Services Task Force (2021) that colonoscopy screening every ten years starting at age 45 and until the age of 75 will have the benefit of 337 life-years gained per 1000 individuals, compared to no screenings. Improving the colonoscopy uptake among the noncompliant population would have an even more significant benefit because noncompliers of cancer screenings tend to exhibit riskier characteristics; for example, they are less likely to undertake preventive care or flu shots (Einav et al. 2020). Our results suggest hospital managers could use a similar program to increase patient colonoscopy uptake among riskier groups.

**7.2.4. Generalizing ML-guided Cancer Screening Outreach** Our findings indicate that ML-guided cancer screening outreach programs could be impactful in real operations by increasing

screening participation and decreasing mortality, in addition to higher detection rates. Similar ML-guided screening outreach programs could be applied in other hospitals and a broader setting (e.g., state- or national-level) or for other diseases (e.g., lung cancer screening). Our findings suggest that managers and policymakers should consider the following factors when implementing ML-guided cancer screening protocols:

- **Setting the threshold:** A key decision of utilizing an ML-guided outreach policy is setting the threshold risk score for treatment. Our partner hospital set the threshold by considering the service capacity (the next bullet point). Lowering the threshold increases the number of patients invited for colonoscopy, which requires additional staffing for outreach and screening. To apply ML-guided programs in a broader setting, more research should be conducted analyzing the cost and benefits to patients, together with the performance of the ML, by different thresholds, because the program’s local/global treatment effects will depend on the thresholds.

- **Service capacity:** When designing the screening program, the service capacity must be considered because the program effect could depend on the service elements, including the outreach, screening, and subsequent treatment capacity. [Boutilier et al. \(2024\)](#) expound and underscore the importance of capacity for interventions in service settings.

- **Mode of communication:** Geisinger made a personalized outreach by a phone call, and in case the patient did not answer the call, an official letter was sent, as described in [Section 3](#). The mode of communication could be a critical factor for an effective patient outreach program ([Phillips et al. 2015](#)). Additionally, the mode of communication has cost implications – for example, staffing costs associated with nurse outreach via phone calls. Our partner hospital allocated 16 hours weekly for a nurse coordinator to perform these outreach phone calls.

- **Fatality of a disease:** CRC is the second most deadly cancer in the U.S., and due to this fatality, we find that the ML-guided screening program decreases mortality. The effect on mortality may not be generalizable to other diseases that are less lethal.

- **Ease of screening:** We find that our program increased screening participation by 6 pp, a substantial jump for CRC screening, but relative to other cancers, this magnitude might be small. This is because colonoscopy is an invasive and discomfoting test, and that is why CRC screening rates are below those seen for breast or prostate cancer ([Roy et al. 2006](#)). For other types of cancers with lower screening barriers, the ML-guided screening policy may have a bigger impact than we find here.

### 7.3. Limitations

Our paper has a few limitations, primarily related to the limitations of our data. First, because our data does not provide the cause of the patient’s death, we cannot analyze the effects on CRC-related deaths. Instead, our analysis of mortality pertains to all-cause deaths. As such, we expect

that some of the reduction is likely due to CRC-related deaths as well as spillover effects on the care of diseases other than CRC; unfortunately, we are not able to tease this out. Information on the cause of death could also enable inference about missed CRC diagnosis. Future research could analyze the effects of a similar program on CRC or cancer-related deaths.

Next, although detecting and removing advanced adenomas during colonoscopy is shown to be effective in reducing CRC incidence rates and mortality (Winawer et al. 1993, Zauber et al. 2012), we cannot prove with certainty that this is a mechanism of reduced mortality in our finding as we do not have data on the detection of advanced adenomas. Future research could collect more detailed colonoscopy procedure data to explain the benefits of detecting advanced adenomas in a similar program.

Third, because our data covers only until 2023, we are able to analyze mortality rates two years after the risk score calculated date. We find that the effects on mortality become more significant over time, as we find more negative estimates in 2-year mortality than in 1-year mortality. The effect of ML-guided cancer screenings may be more significant in the long run. Future research could track the participants for longer to understand the long-term effects of ML-guided cancer screenings.

Fourth, our research could not analyze the heterogeneous effects of the ML-guided cancer screening outreach by the mode of communication because the communication channel (phone call vs. written letter) was not randomized, and only about 12% of the flagged patients were communicated via a letter after no response to three phone calls. It would be important to analyze the effects of ML-guided cancer screening programs through different outreach channels to find the most effective communication channels.

Lastly, our analysis focuses on the intervention at one health system, which began before the U.S. Preventive Services Task Force (2021) lowered the screening age to 45 from 50 in 2021. Should a patient population in a different health system be significantly different from Geisinger’s (e.g., differences in PCP visit cadence and CBC tests), the impact of an ML-guided screening outreach program could be different than what this paper finds. Future research could generalize our findings by studying similar interventions at other institutions and including adults aged 45 and older per the updated screening guidelines.

## References

- Adams A, Kluender R, Mahoney N, Wang J, Wong F, Yin W (2022) The impact of financial assistance programs on health care utilization: Evidence from Kaiser Permanente. *American Economic Review: Insights*, 4(3):389–407.
- Adjerid I, Ayvaci MU, Özer Ö (2023) Value of algorithm-enabled process innovation: The case of sepsis. *Manufacturing Service Oper. Management* 25(4):1545—1566.

- Almond D, Doyle Jr. JJ (2011) After midnight: A regression discontinuity design in length of postpartum hospital stays. *American Economic Journal: Economic Policy* 3(3):1–34.
- American Cancer Society (2023) Colorectal cancer facts & figures 2023–2025. Report, Atlanta: American Cancer Society.
- American Cancer Society National Colorectal Cancer Roundtable (2024) 80% in Every Community. <https://nccrt.org/our-impact/80-in-every-community/>.
- Basch CE, Wolf RL, Brouse CH, Shmukler C, Neugut A, DeCarlo LT, Shea S (2006) Telephone outreach to increase colorectal cancer screening in an urban minority population. *American Journal of Public Health* 96(12):2246–2253.
- Bharadwaj P, Løken KV, Neilson C (2013) Early life health interventions and academic achievement. *Amer. Econom. Rev.* 103(5):1862–1891.
- Bertsimas D, Pauphilet J (2023) Hospital-wide inpatient flow optimization. *Management Sci.* Articles in Advance, 1–19.
- Birks J, Bankhead C, Holt TA, Fuller A, Patnick J (2017) Evaluation of a prediction model for colorectal cancer: Retrospective analysis of 2.5 million patient records. *Cancer Medicine* 6(10):2453–2460.
- Boutilier JJ, Jónasson JO, Yoeli E (2022) Improving tuberculosis treatment adherence support: The case for targeted behavioral interventions. *Manufacturing Service Oper. Management* 24(6):2925–2943.
- Boutilier J, Jonasson JO, Li H, Yoeli E (2024) Randomized controlled trials of service interventions: The impact of capacity constraints. *arXiv preprint arXiv:2407.21322*.
- Brenner H, Stock C, Hoffmeister M (2014) Effect of screening sigmoidoscopy and screening colonoscopy on colorectal cancer incidence and mortality: systematic review and meta-analysis of randomised controlled trials and observational studies. *BMJ* 348:g2467.
- Bretthauer M, Kaminski MF, Løberg M, Zauber AG, Regula J, Kuipers EJ, ..., Adami HO (2016) Population-based colonoscopy screening for colorectal cancer: A randomized clinical trial. *JAMA Internal Medicine* 176(7):894–902.
- Bretthauer M, Løberg M, Wieszczy P, Kalager M, Emilsson L, Garborg K, ..., Kaminski MF (2022) Effect of colonoscopy screening on risks of colorectal cancer and related death. *New England Journal of Medicine* 387(17):1547–1556.
- Bundorf MK, Polyakova M, Tai-Seale M (2024) How do consumers interact with digital expert advice? Experimental evidence from health insurance. *Management Sci.* 70(11):7617–7643.
- Calonico S, Cattaneo MD, Titiunik R (2014) Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6):2295–2326.
- Calonico S, Cattaneo MD, Titiunik R (2015) Rdrobust: An R package for robust nonparametric inference in regression-discontinuity designs. *R Journal* 7(1):38–51.

- Calonico S, Jawadekar N, Kezios K, Al Hazzouri AZ (2024) Regression discontinuity design studies: a guide for health researchers. *BMJ*, 384:e072254.
- Card D, Dobkin C, Maestas N (2009) Does Medicare save lives? *The Quarterly Journal of Economics* 124(2):597–636.
- Castells A, Quintero E (2015) Programmatic screening for colorectal cancer: The COLONPREV study. *Digestive Diseases and Sciences* 60:672–680.
- Cattaneo MD, Idrobo N, Titiunik R (2020) A Practical Introduction to Regression Discontinuity Designs: Foundations. Cambridge University Press.
- Cattaneo MD, Titiunik R (2022) Regression discontinuity designs. *Annual Review of Economics* 14:821–851.
- Chan TC, Mahmood R, O’Connor DL, Stone D, Unger S, Wong RK, Zhu IY (2025) Got (optimal) milk? Pooling donations in human milk banks with machine learning and optimization. *Manufacturing Service Oper. Management* 27(6), 1721-1739.
- Clark D, Royer H (2013) The effect of education on adult mortality and health: Evidence from Britain. *Amer. Econom. Rev.* 103(6):2087–2120.
- DeGross A, Gressard L, Glover-Kudon R, Rice K, Tharpe FS, Escoffery C, Gersten J, Butterly L (2019) Assessing the implementation of a patient navigation intervention for colonoscopy screening. *BMC Health Services Research* 19(803):1–11.
- Dominitz JA, Robertson DJ, Ahnen DJ, Allison JE, Antonelli M, Boardman KD, ..., CONFIRM Study Group (2017) Colonoscopy vs. fecal immunochemical test in reducing mortality from colorectal cancer (CONFIRM): Rationale for study design. *The American Journal of Gastroenterology* 112(11):1736–1746.
- Einav L, Finkelstein A, Ostrom T, Ostriker A, Williams H (2020) Screening and selection: The case of mammograms. *Amer. Econom. Rev.* 110(12):3836–3870.
- Erenay FS, Alagoz O, Said A (2014) Optimizing colonoscopy screening for colorectal cancer prevention and surveillance. *Manufacturing Service Oper. Management* 16(3):381–400.
- European Commission (2022) European health union: Commission welcomes adoption of new EU cancer screening recommendations. (December 9). Accessed June 19, 2024, [https://ec.europa.eu/commission/presscorner/detail/en/ip\\_22\\_7548](https://ec.europa.eu/commission/presscorner/detail/en/ip_22_7548).
- Fisher DA, Princic N, Miller-Wilson LA, Wilson K, Limburg P (2022) Healthcare costs of colorectal cancer screening and events following colonoscopy among commercially insured average-risk adults in the United States. *Current Medical Research and Opinion* 38(3):427–434.
- Forsberg A, Westerberg M, Metcalfe C, Steele R, Blom J, Engstrand L, ..., Hultcrantz R (2022) Once-only colonoscopy or two rounds of faecal immunochemical testing 2 years apart for colorectal cancer screening (SCREESCO): Preliminary report of a randomised controlled trial. *The Lancet Gastroenterology & Hepatology* 7(6):513–521.

- Gao SY, He Y, Zhang R, Zheng Z, Wei SLS (2025) Optimizing initial screening for colorectal cancer detection with adherence behavior. *Management Sci.* 71(9):7516–7536.
- Gelman A, Imbens G (2019) Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics* 37(3):447–456.
- Gould MK, Huang BZ, Tammemagi MC, Kinar Y, Shiff R (2021) Machine learning for early lung cancer identification using routine clinical and laboratory data. *American Journal of Respiratory and Critical Care Medicine* 204(4):445–453.
- Gupta S, Halm EA, Rockey DC, Hammons M, Koch M, Carter, E, Valdez L et al. (2013) Comparative effectiveness of fecal immunochemical test outreach, colonoscopy outreach, and usual care for boosting colorectal cancer screening among the underserved: a randomized clinical trial. *JAMA Internal Medicine* 173(18):1725–1732.
- Hahn J, Todd P, Van der Klaauw W (2001) Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1):201–209.
- Honeycutt S, Green R, Ballard D, Hermstad A, Brueder A, Haardörfer R, Yam J, Arriola KJ (2013) Evaluation of a patient navigation program to promote colorectal cancer screening in rural Georgia, USA. *Cancer* 119(16): 3059–3066.
- Hornbrook MC, Goshen R, Choman E, O’Keeffe-Rosetti M, Kinar Y, Liles EG, Rust KC (2017) Early colorectal cancer detected by machine learning model using gender, age, and complete blood count data. *Digestive Diseases and Sciences* 62:2719–2727.
- Hu Y, Chan CW, Dong J (2025) Prediction-driven surge planning with application in the emergency department. *Management Sci.* 71(3):2079–2126.
- Imbens G, Kalyanaraman K (2012) Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies* 79(3):933–959.
- Jandorf L, Gutierrez Y, Lopez J, Christie J, Itzkowitz SH (2005) Use of a patient navigator to increase colorectal cancer screening in an urban neighborhood health clinic. *Journal of Urban Health* 82(2):216–224.
- Jordan MI, Mitchell TM (2015) Machine learning: Trends, perspectives, and prospects. *Science* 349(6245):255–260.
- Kahi CJ, Imperiale TF, Juliar BE, Rex DK (2009) Effect of screening colonoscopy on colorectal cancer incidence and mortality. *Clinical Gastroenterology and Hepatology* 7(7):770–775.
- Kinar Y, Kalkstein N, Akiva P, Levin B, Half EE, Goldshtein I, et al. (2016) Development and validation of a predictive model for detection of colorectal cancer in primary care by analysis of complete blood counts: a binational retrospective study. *Journal of the American Medical Informatics Association* 23(5):879–890.

- Kinar, Y, Akiva P, Choman E, Kariv R, Shalev V, Levin B, et al. (2017) Performance analysis of a machine learning flagging system used to identify a group of individuals at a high risk for colorectal cancer. *PloS one* 12(2):e0171759.
- Kraus M, Feuerriegel S, Saar-Tsechansky M (2024) Data-driven allocation of preventive care with application to diabetes mellitus type II. *Manufacturing Service Oper. Management* 26(1):137–153.
- Lee YH, Kung PT, Wang YH, Kuo WY, Kao SL, Tsai WC (2019) Effect of length of time from diagnosis to treatment on colorectal cancer survival: A population-based study. *PloS one* 14(1): e0210465.
- Lee DS, Lemieux T (2010) Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2):281-355.
- Leone LA, Reuland DS, Lewis CL, Ingle M, Erman B, Summers TJ, DuBard CA, Pignone MP (2013) Reach, usage, and effectiveness of a Medicaid patient navigator intervention to increase colorectal cancer screening, Cape Fear, North Carolina, 2011. *Prev. Chronic Dis.* 10:120272.
- Leshno M, Halpern Z, Arber N (2003) Cost-effectiveness of colorectal cancer screening in the average risk population. *Health Care Management Sci.* 6:165–174.
- Lieberman DA, Weiss DG, Bond JH, Ahnen DJ, Garewal H, Harford WV, ..., Chejfec G (2000) Use of colonoscopy to screen asymptomatic adults for colorectal cancer. *New England Journal of Medicine*, 343(3):162–168.
- Lin W, Kim SH, Tong J (2022) What drives algorithm use? An empirical analysis of algorithm use in type 1 diabetes self-management. Working paper. Available at SSRN: <https://ssrn.com/abstract=3891832> or <http://dx.doi.org/10.2139/ssrn.3891832>.
- McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2):698–714.
- National Cancer Institute (2024) Cancer Stat Facts: Colorectal cancer. Accessed March 26, 2024, <https://seer.cancer.gov/statfacts/html/colorect.html>.
- Percac-Lima S, Grant RW, Green AR, Ashburner JM, Gamba G, Oo S, Richter JM, Atlas SJ (2009) A culturally tailored navigator program for colorectal cancer screening in a community health center: a randomized, controlled trial. *J. Gen. Intern. Med.* 24(2):211–217.
- Phillips L, Hendren S, Humiston S, Winters P, Fiscella K (2015) Improving breast and colon cancer screening rates: a comparison of letters, automated phone calls, or both. *The Journal of the American Board of Family Medicine* 28(1):46–54.
- Quintero E, Castells A, Bujanda L, Cubiella J, Salas D, Lanas Á, ..., González-Navarro A (2012) Colonoscopy versus fecal immunochemical testing in colorectal-cancer screening. *New England Journal of Medicine* 366(8):697–706.
- Rice K, Gressard L, DeGroff A, Gersten J, Robie J, Leadbetter S, ..., Butterly L (2017) Increasing colonoscopy screening in disparate populations: Results from an evaluation of patient navigation in the New Hampshire Colorectal Cancer Screening Program. *Cancer* 123(17):3356–3366.

- Roy HK, Backman V, Goldberg MJ (2006) Colon cancer screening: The good, the bad, and the ugly. *Archives of Internal Medicine* 166(20):2177–2179.
- Schneider JL, Layefsky E, Udaltsova N, Levin TR, Corley DA (2020) Validation of an algorithm to identify patients at risk for colorectal cancer based on laboratory test and demographic data in diverse, community-based population. *Clinical Gastroenterology and Hepatology* 18(12):2734–2741.
- Sekhon JS, Titiunik R (2017) On interpreting the regression discontinuity design as a local experiment. In *Regression discontinuity designs: Theory and applications. Advances in Econometrics* 38:1–28. Emerald Publishing Limited.
- Shaukat A, Levin TR (2022) Current and future colorectal cancer screening strategies. *Nature reviews: Gastroenterology & Hepatology* 19(8):521–531.
- Singal AG, Gupta S, Skinner CS, Ahn C, Santini NO, Agrawal D, Mayorga CA et al. (2017) Effect of colonoscopy outreach vs fecal immunochemical test outreach on colorectal cancer screening completion: a randomized clinical trial. *JAMA* 318(9):806–815.
- Stark GF, Hart GR, Nartowt BJ, Deng J (2019) Predicting breast cancer risk using personal health data and machine learning models. *Plos one* 14(12):e0226765.
- Thistlethwaite DL, Campbell DT (1960) Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology* 51(6):309.
- Tørring ML, Frydenberg M, Hansen RP, Olesen F, Vedsted P (2013) Evidence of increasing mortality with longer diagnostic intervals for five common cancers: a cohort study in primary care. *European Journal of Cancer* 49(9):2187–2198.
- Underberger D, Boell K, Orr J, Siegrist C, Hunt S (2022) Collaboration to improve colorectal cancer screening using machine learning. *NEJM Catalyst Innovations in Care Delivery* 3(4):1–10.
- U.S. Preventive Services Task Force (2016) Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *JAMA* 315(23):2564–2575.
- U.S. Preventive Services Task Force (2021) Screening for colorectal cancer: US Preventive Services Task Force recommendation statement. *JAMA* 325(19):1965–1977.
- Winawer SJ, Zauber AG, Ho MN, O’Brien MJ, Gottlieb LS, Sternberg SS, ..., National Polyp Study Workgroup. (1993) Prevention of colorectal cancer by colonoscopic polypectomy. *New England Journal of Medicine* 329(27), 1977-1981.
- Zauber AG, Winawer SJ, O’Brien MJ, Lansdorp-Vogelaar I, van Ballegooijen M, Hankey BF, ..., Waye JD (2012) Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths. *New England Journal of Medicine* 366(8):687–696.

## Online Appendices

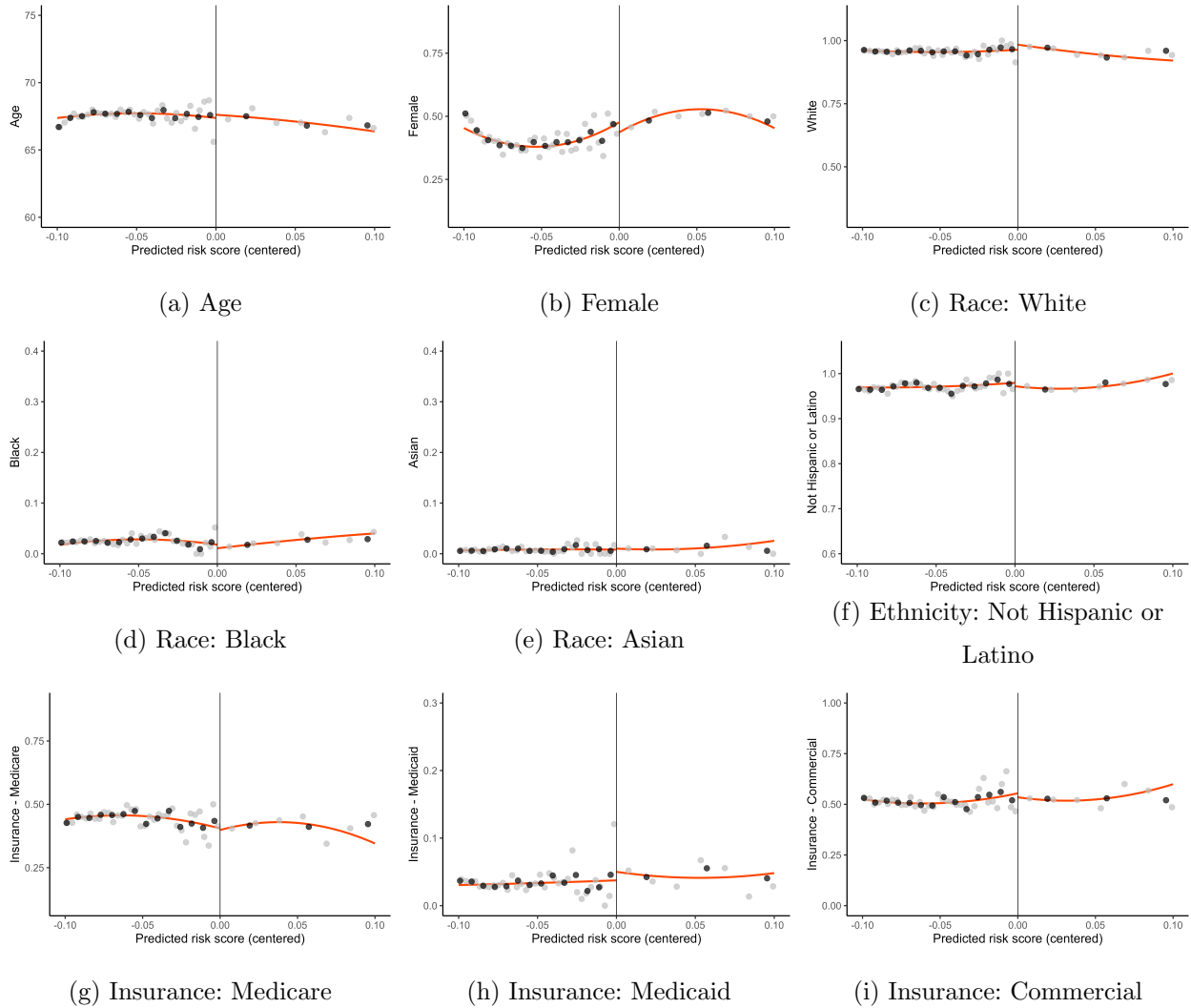
### Online Appendix A: Sampling Steps

**Table OA.1 Detailed Sampling Steps**

	Sampling step	Sample size after each sampling step
1	Risk scores collected from July 2019 to December 2022	72,369
2	Exclude flagged risk scores of patients who had a colonoscopy before the outreach	72,364
3	Exclude flagged risk scores that were not outreached after a chart review (because, for example, the patient was deceased or admitted to the hospital)	72,094
4	Exclude risk scores of patients who are aged 50	69,482
5	For patients with multiple risk scores, limit one risk score per each month per patient (include only the patient's highest risk score in each month)	62,485 Treated: 1,521 Control: 60,964

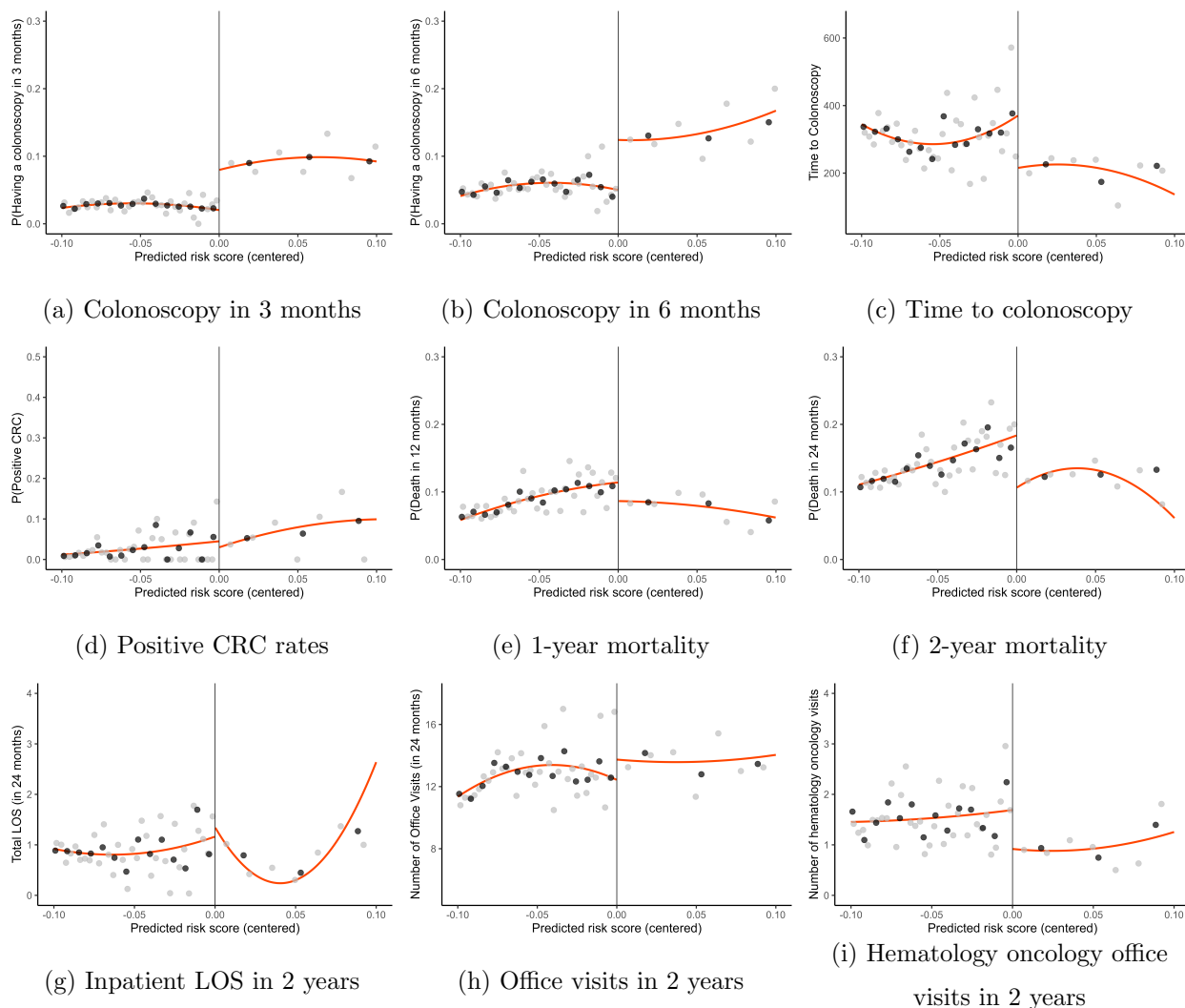
## Online Appendix B: Figures with Quadratic Regressions

**Figure OA.1** Balance of Predetermined Characteristics with Quadratic Regressions



*Notes.* The figure plots the characteristics of the sample. Black points present the average value of the specified variable for risk scores with 20 bins on each side of the cutoff. Gray points present the average value of the specified variable for risk scores with 50 bins on each side of the cutoff. The solid vertical line represents the cutoff risk score value. The solid, red trend lines are the predicted values from local quadratic regression fits of the specified variable on the score, fitted separately above and below the cutoff, that uses a triangular kernel and a bandwidth of 0.10 on each side.

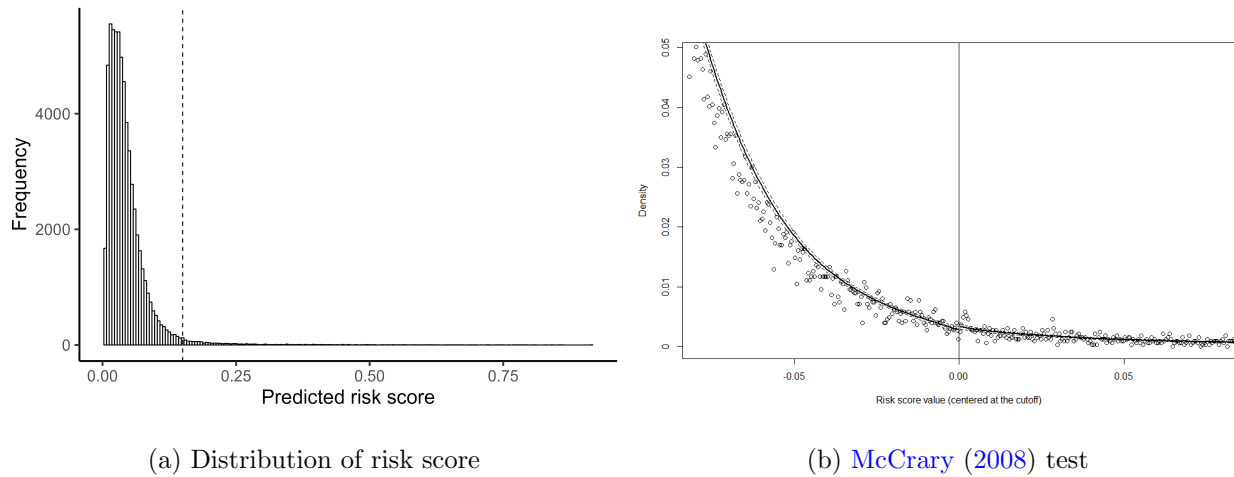
Figure OA.2 RD Plots with Quadratic Regressions



Notes. Black points present the average value of the specified variable for risk scores with 20 bins on each side of the cutoff. Gray points present the average value of the specified variable for risk scores with 50 bins on each side of the cutoff. The solid vertical line represents the cutoff risk score value. The solid, red trend lines are the predicted values from local quadratic regression fits of the specified variable on the score, fitted separately above and below the cutoff, that uses a triangular kernel and a bandwidth of 0.10 on each side. CRC: colorectal cancer, LOS: length of stay.

## Online Appendix C: Distribution of Risk Score

**Figure OA.3** Density of Running Variable



*Notes.* Panel (a) shows the distribution of predicted risk score for the entire range, with a bin width of 0.005. Panel (b) shows the densities computed over bins with a width of 0.0005, around the cutoff point. The line plots a nonparametric regression to each side of the distribution following [McCrary \(2008\)](#), testing for a discontinuity at zero. The test fails to reject the null hypothesis of no discontinuity in the density at the cutoff ( $p$ -value = 0.154).

**Online Appendix D: Analysis of Inpatient LOS by Specialty****Table OA.2 Effects on Inpatient LOS by Specialty at the Cutoff**

Outcome Specialty	2-year LOS			
	ICU (1)	Medical/Surgical (2)	Oncology (3)	Radiology (4)
<i>Panel A. Local linear regression</i>				
Flag	-0.177 (0.174)	0.125 (0.875)	-0.075 (0.065)	0.003 (0.002)
Controls	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.045, 0.032]	[-0.062, 0.073]	[-0.077, 0.096]	[-0.077, 0.096]
Mean below cutoff	0.111	0.515	0.180	0.004
Observations	682	1,246	2,016	2,016
<i>Panel B. Local quadratic regression</i>				
Flag	0.036 (0.185)	0.197 (1.038)	-0.006 (0.061)	0.001 (0.003)
Controls	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.034, 0.061]	[-0.081, 0.120]	[-0.100, 0.145]	[-0.100, 0.145]
Mean below cutoff	0.157	0.484	0.163	0.005
Observations	535	2,340	4,523	4,523

*Notes.* Panels A and B present the estimates using local linear regression and local quadratic regression, respectively, with a triangular weighting kernel. Control variables include age, sex, and race. Bandwidth shows the interval centered around the cutoff point, 0.150, using MSE optimal bandwidths (Calonico et al. 2014). For columns 3 and 4, MSE optimal bandwidths cannot be calculated due to too many zero values; we use the bandwidths in Table 4 column 5 and column 6. The observations row shows effective observations within the bandwidth. Robust standard errors are reported in parentheses. Total observations: 15,633 include Geisinger Health Plan holders with risk scores collected until December 2021. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

**Online Appendix E: Analysis of Office Visits by Specialty****Table OA.3 Effects on 2-year Office Visits by Relevant Specialty at the Cutoff**

Outcome Specialty	Number of visits in 2 years			
	Family Medicine	General Surgery	Internal Medicine	Radiation Oncology
	(1)	(2)	(3)	(4)
<i>Panel A. Local linear regression</i>				
Flag	0.245 (0.620)	0.286 (0.153)	-0.006 (0.183)	0.025 (0.128)
Controls	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.075, 0.097]	[-0.041, 0.085]	[-0.043, 0.146]	[-0.036, 0.085]
Mean below cutoff	4.485	0.158	0.306	0.146
Observations	1,885	710	806	621
<i>Panel B. Local quadratic regression</i>				
Flag	0.235 (0.724)	0.281 (0.174)	-0.005 (0.220)	0.067 (0.151)
Controls	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.087, 0.159]	[-0.052, 0.159]	[-0.070, 0.204]	[-0.055, 0.105]
Mean below cutoff	4.466	0.207	0.290	0.153
Observations	2,861	1,039	1,699	1,065

*Notes.* Panels A and B present the estimates using local linear regression and local quadratic regression, respectively, with a triangular weighting kernel. Control variables include age, sex, and race. Bandwidth shows the interval centered around the cutoff point, 0.150, using MSE optimal bandwidths (Calonico et al. 2014). The observations row shows effective observations within the bandwidth. Robust standard errors are reported in parentheses. Total observations: 15,633 include Geisinger Health Plan holders with risk scores collected until December 2021. \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

**Online Appendix F: Placebo Outcomes**

**Table OA.4 Effects on Unrelated Outcomes**

Outcome	2-year LOS: Orthopedics		2-year LOS: Psychiatry	
	Linear (1)	Quadratic (2)	Linear (3)	Quadratic (4)
Flag	0.001 (0.019)	-0.010 (0.026)	-0.017 (0.016)	-0.021 (0.021)
Controls	Yes	Yes	Yes	Yes
Mean below cutoff	0.003	0.003	0.018	0.010
Observations	1,091	2,152	1,091	2,152

*Notes.* The table presents the estimates using local linear regression (columns 1 and 3) and local quadratic regression (columns 2 and 4). Control variables include age, sex, and race. Robust standard errors are reported in parentheses. We use the bandwidths in [Table 5](#) column 1 and column 2 because MSE optimal bandwidths cannot be calculated due to too many zero values. As [Table 5](#), analyses include Geisinger Health Plan holders with risk scores collected until December 2021 (Total observations: 15,633). <sup>+</sup> $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

**Online Appendix G: Robustness to Different Kernels****Table OA.5** Effects on Colonoscopy Uptake and Mortality at the Cutoff Using Different Kernels

Outcome Kernels	Colonoscopy in 3 months				2-year mortality			
	Uniform		Epanechnikov		Uniform		Epanechnikov	
	(1) Linear	(2) Quadratic	(3) Linear	(4) Quadratic	(5) Linear	(6) Quadratic	(7) Linear	(8) Quadratic
Flag	0.053** (0.019)	0.050* (0.024)	0.063** (0.019)	0.061** (0.021)	-0.087** (0.033)	-0.080* (0.035)	-0.068* (0.028)	-0.075* (0.035)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.042, 0.070]	[-0.050, 0.093]	[-0.068, 0.063]	[-0.090, 0.108]	[-0.053, 0.039]	[-0.086, 0.093]	[-0.080, 0.068]	[-0.107, 0.086]
Mean below cutoff	0.025	0.029	0.029	0.029	0.150	0.138	0.140	0.121
Observations	2,622	3,449	5,803	13,072	2,753	8,520	6,958	17,759
Control	1,834	2,549	5,058	12,103	2,304	7,824	6,356	17,086
Treated	788	900	745	969	449	696	602	673

*Notes.* The table presents the estimate using uniform and Epanechnikov kernel functions. Control variables include age, sex, race, and insurance type (Medicare, Medicaid, military, commercial, self-pay). Bandwidth shows the interval centered at the cutoff point, 0.150. The observations row shows effective observations within the bandwidth. Robust standard errors are reported in parentheses. For columns 7 to 8, risk scores collected before January 2022 are included. <sup>+</sup> $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

**Online Appendix H: Robustness to Different Samples****Table OA.6** Effects on Colonoscopy Uptake and Mortality at the Cutoff Using Different Samples

Sample Outcome	Sample including aged 50					
	Colonoscopy in 3 months		Colonoscopy in 6 months		2-year mortality	
	Linear (1)	Quadratic (2)	Linear (3)	Quadratic (4)	Linear (5)	Quadratic (6)
Flag	0.062** (0.019)	0.063** (0.022)	0.073** (0.023)	0.075** (0.025)	-0.070* (0.028)	-0.074* (0.034)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.069, 0.073]	[-0.085, 0.108]	[-0.067, 0.072]	[-0.104, 0.140]	[-0.086, 0.070]	[-0.111, 0.101]
Mean below cutoff	0.030	0.029	0.060	0.053	0.136	0.117
Observations	6,088	11,037	5,788	21,387	8,696	20,700

*Notes.* The table presents the estimates using local linear regression (columns 1, 3, and 5) or local quadratic regression (columns 2, 4, and 6) with a triangular weighting kernel. Bandwidth shows the interval centered around the cutoff point, 0.150, using MSE optimal bandwidths (Calonico et al. 2014). The observations row shows effective observations within the bandwidth. Robust standard errors are reported in parentheses. Total observations: 64,925 (columns 1, 2, 3, and 4), 48,895 (columns 5 and 6, includes risk scores until December 2021).  $^+p < 0.1$ ,  $^*p < 0.05$ ,  $^{**}p < 0.01$ ,  $^{***}p < 0.001$ .

**Table OA.7** Effects on Colonoscopy Uptake and Mortality at the Cutoff Using One Risk Score per Patient

Sample Outcome	One risk score per patient					
	Colonoscopy in 3 months		Colonoscopy in 6 months		2-year mortality	
	Linear (1)	Quadratic (2)	Linear (3)	Quadratic (4)	Linear (5)	Quadratic (6)
Flag	0.069** (0.021)	0.073** (0.025)	0.084** (0.027)	0.085** (0.030)	-0.067* (0.031)	-0.078* (0.040)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.071, 0.079]	[-0.085, 0.129]	[-0.064, 0.077]	[-0.095, 0.162]	[-0.081, 0.067]	[-0.099, 0.086]
Mean below cutoff	0.038	0.041	0.070	0.071	0.116	0.102
Observations	4,116	6,305	3,394	8,423	4,189	7,239

*Notes.* The table presents the estimates using local linear regression (columns 1, 3, and 5) or local quadratic regression (columns 2, 4, and 6) with a triangular weighting kernel. Control variables include age, sex, race, and insurance type (Medicare, Medicaid, military, commercial, self-pay). Bandwidth shows the interval centered around the cutoff point, 0.150, using MSE optimal bandwidths (Calonico et al. 2014). The observations row shows effective observations within the bandwidth. Robust standard errors are reported in parentheses. Total observations: 32,782 (columns 1 and 2), 24,227 (columns 5 and 6, includes risk scores until December 2021).  $^+p < 0.1$ ,  $^*p < 0.05$ ,  $^{**}p < 0.01$ ,  $^{***}p < 0.001$ .

**Online Appendix I: Analysis using Log-transformed Risk Scores****Table OA.8** Effects on Colonoscopy Uptake and Mortality at the Cutoff Using Log-transformed Risk Scores

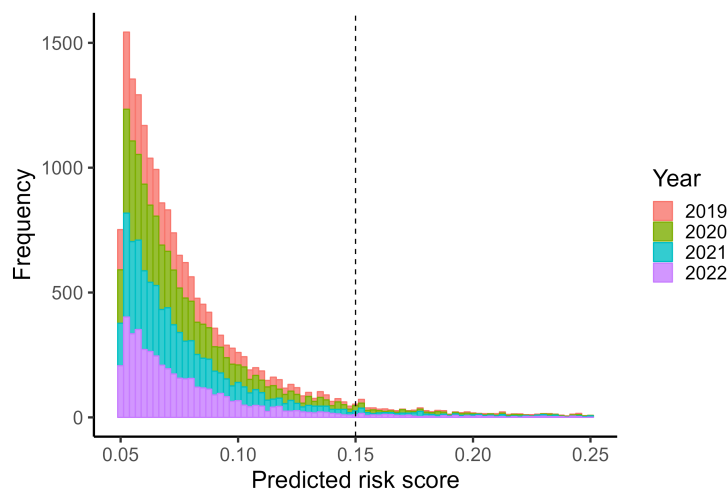
Outcome	Colonoscopy in 3 months		Colonoscopy in 6 months		2-year mortality	
	Linear (1)	Quadratic (2)	Linear (3)	Quadratic (4)	Linear (5)	Quadratic (6)
Flag	0.052** (0.018)	0.052* (0.022)	0.062** (0.022)	0.067* (0.027)	-0.063* (0.026)	-0.070* (0.032)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-1.490, 0.432]	[-1.579, 0.508]	[-1.494, 0.394]	[-1.296, 0.527]	[-1.574, 0.361]	[-2.047, 0.466]
Mean below cutoff	0.033	0.033	0.060	0.060	0.079	0.079
Observations	31,857	34,938	32,017	25,246	26,308	35,850

*Notes.* The table presents the estimates using local linear regression (columns 1, 3, and 5) or local quadratic regression (columns 2, 4, and 6) with a triangular weighting kernel. Control variables include age, sex, race, and insurance type (Medicare, Medicaid, military, commercial, self-pay). Bandwidth shows the interval of log-transformed risk score centered around the cutoff point,  $\log(0.150)$ , using MSE optimal bandwidths (Calonic et al. 2014). The observations row shows effective observations within the bandwidth. Robust standard errors are reported in parentheses. Total observations: 62,485 (columns 1, 2, 3, and 4); 46,994 (columns 5 and 6). <sup>+</sup> $p < 0.1$ , \*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$ .

**Online Appendix J: Analysis using Clustered Standard Errors****Table OA.9** Effects on Colonoscopy Uptake and Mortality at the Cutoff Using Clustered Standard Errors

Outcome	Colonoscopy in 3 months		Colonoscopy in 6 months		2-year mortality	
	Linear (1)	Quadratic (2)	Linear (3)	Quadratic (4)	Linear (5)	Quadratic (6)
Flag	0.060** (0.018)	0.059** (0.021)	0.069** (0.023)	0.074** (0.026)	-0.062* (0.030)	-0.071* (0.036)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.069, 0.074]	[-0.086, 0.112]	[-0.071, 0.077]	[-0.107, 0.112]	[-0.074, 0.097]	[-0.097, 0.147]
Mean below cutoff	0.028	0.029	0.058	0.052	0.147	0.129
Observations	6,116	11,452	6,545	23,996	5,679	12,989

*Notes.* The table presents the estimates using local linear regression (columns 1, 3, and 5) or local quadratic regression (columns 2, 4, and 6) with a triangular weighting kernel. Control variables include age, sex, race, and insurance type (Medicare, Medicaid, military, commercial, self-pay). Bandwidth shows the interval of risk score centered around the cutoff point,  $\log(0.150)$ , using MSE optimal bandwidths (Calonico et al. 2014). The observations row shows effective observations within the bandwidth. Cluster (patient)-robust standard errors are reported in parentheses. Total observations: 62,485 (columns 1, 2, 3, and 4); 46,994 (columns 5 and 6). <sup>+</sup> $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

**Online Appendix K: Robustness to COVID-19 Pandemic****Figure OA.4 Density of Running Variable**

(a) Distribution of risk score by year

*Notes.* Figure plots the distribution of predicted risk score around the cutoff by year (color), with a bin width of 0.005.

**Table OA.10 Analysis with Year Effects**

Outcome	With year effects					
	Colonoscopy in 3 months		Colonoscopy in 6 months		2-year mortality	
	Linear (1)	Quadratic (2)	Linear (3)	Quadratic (4)	Linear (5)	Quadratic (6)
Flag	0.069** (0.021)	0.073** (0.025)	0.084** (0.027)	0.085** (0.030)	-0.067* (0.031)	-0.078* (0.040)
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.071, 0.079]	[-0.085, 0.129]	[-0.064, 0.077]	[-0.095, 0.162]	[-0.081, 0.067]	[-0.099, 0.086]
Mean below cutoff	0.038	0.041	0.070	0.071	0.116	0.102
Observations	4,116	6,305	3,394	8,423	4,189	7,239

*Notes.* The table presents the estimates using local linear regression (columns 1, 3, and 5) or local quadratic regression (columns 2, 4, and 6) with a triangular weighting kernel. Control variables include age, sex, race, insurance type (Medicare, Medicaid, military, commercial, self-pay), and year indicators. Bandwidth shows the interval centered around the cutoff point, 0.150, using MSE optimal bandwidths (Calonico et al. 2014). The observations row shows effective observations within the bandwidth. Robust standard errors are reported in parentheses. Total observations: 32,782 (columns 1 and 2), 24,227 (columns 5 and 6, includes risk scores until December 2021). <sup>+</sup> $p < 0.1$ , \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

**Table OA.11 Analysis with Year Effects and Excluding Deaths in 2020**

Sample Outcome	Excluding deaths in 2020			
	Colonoscopy in 3 months		2-year mortality	
	Linear (1)	Quadratic (2)	Linear (3)	Quadratic (4)
Flag	0.062** (0.019)	0.061** (0.022)	-0.068* (0.028)	-0.075* (0.032)
Controls	Yes	Yes	Yes	Yes
Bandwidth (interval)	[-0.068, 0.069]	[-0.086, 0.105]	[-0.080, 0.055]	[-0.103, 0.102]
Mean below cutoff	0.030	0.030	0.108	0.093
Observations	5,774	11,170	6,601	15,354

*Notes.* The table presents the estimates using local linear regression (columns 1 and 3) or local quadratic regression (columns 2 and 4) with a triangular weighting kernel. Control variables include age, sex, race, and insurance type (Medicare, Medicaid, military, commercial, self-pay) in all columns and year dummies in columns 1 and 2. Bandwidth shows the interval centered around the cutoff point, 0.150, using MSE optimal bandwidths (Calonico et al. 2014). Robust standard errors are reported in parentheses. Total observations: 61,478 (columns 1 and 2), and 45,987 (columns 3 and 4, includes risk scores until December 2021).

## Online Appendix L: Analysis of CRC Diagnosis Stages and Time to Diagnosis

In this section, we explore a potential mechanism underlying reduced mortality by analyzing the time to CRC diagnosis and the stages at diagnosis. It is important to interpret the findings here with caution, given the limited number of diagnosed cases: 59 in the control group and 23 in the treated group. Due to this small sample size, we cannot apply an RD design to estimate the treatment effect on diagnosis stages reliably. 82 observations across the risk scores are insufficient for RD analysis, which requires a larger sample to estimate the regression functions validly.

Additionally, the program's intervention (flagging and inviting patients for colonoscopy) increased colonoscopy uptake among treated group patients but not among control group patients. Consequently, control group patients are more likely to have missed CRC diagnoses. This introduces a potential selection bias when comparing diagnosis stages across groups. For example, control group patients who are not invited for a colonoscopy but are diagnosed with CRC may have more regular medical check-ups, which could influence their diagnosis stage independently of the intervention.

Considering these limitations, we compared the average diagnosis stages between groups. We collected data on CRC stages. Stages are recorded from stage 0 to stage 4 (stage 0 being the earliest stage). In the control group ( $n = 59$ ), the mean diagnosis stage is 2.57, whereas in the treated group ( $n = 23$ ), the mean diagnosis stage is 3.15 ( $t$ -test = 1.99,  $p$ -value = 0.053). This is an expected finding because the treated group has riskier characteristics (e.g., older), which increases the likelihood of developing CRC at more advanced stages. Although the treated group had, on average, later-stage diagnoses than the control group, they exhibited a shorter time to diagnosis from the date of risk score calculation: 211 days for the treated group versus 266 days for the control group ( $t$ -test = -0.903,  $p$ -value = 0.371).

When restricting the sample to within the MSE-optimal bandwidth around the cutoff  $[-0.088, 0.112]$ , this is the bandwidth of Table 3 column (2) in the paper), the average stages are 2.59 for the control group ( $n = 24$ ), and 3.20 for the treated group ( $n = 13$ ,  $t$ -test = 1.689,  $p$ -value = 0.103). The time to diagnosis within this bandwidth is 224 days for the treated group and 240 days for the control group ( $t$ -test = -0.180,  $p$ -value = 0.858). In summary, this exploratory analysis does not find strong evidence that the treated group had significantly earlier CRC diagnoses. However, the findings here should be interpreted with caution, primarily due to the limited number of diagnosed cases and the possibility that control group patients may have missed diagnoses because they were not invited for a colonoscopy by the program. Therefore, there could be more control group patients who develop later-stage CRC without being diagnosed within the sample period, or even die without a CRC diagnosis (for example, they may be diagnosed post-mortem with carcinoma of unknown primary after metastasis). We recommend that future research, with a larger sample size of diagnosed cases, rigorously examine whether earlier diagnosis serves as a mechanism for reducing mortality.

## References

- Calonico S, Cattaneo MD, Titiunik R (2014) Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica* 82(6):2295–2326.
- McCrary J (2008) Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics* 142(2):698–714.