Improving Cancer Outreach Effectiveness Through Targeting and Economic Assessments: Insights from a Randomized Field Experiment

Yixing Chen, Ju-Yeon Lee, Shrihari (Hari) Sridhar, Vikas Mittal, Katharine McCallister, and Amit G. Singal

Abstract

Patients at risk for hepatocellular carcinoma or liver cancer should undergo semiannual screening tests to facilitate early detection, effective treatment options at lower cost, better recovery prognosis, and higher life expectancy. Health care institutions invest in direct-to-patient outreach marketing to encourage regular screening. They ask the following questions: (1) Does the effectiveness of outreach vary among patients and over time?; (2) What is the return on outreach?; and (3) Can patient-level targeted outreach increase the return? The authors use a multiperiod, randomized field experiment involving 1,800 patients. Overall, relative to the usual-care condition, outreach alone (outreach with patient navigation) increases screening completion rates by 10–20 (13–24) percentage points. Causal forests demonstrate that patient-level treatment effects vary substantially across periods and by patients' demographics, health status, visit history, health system accessibility, and neighborhood socioeconomic status, thereby facilitating the implementation of the targeted outreach program. A simulation shows that the targeted outreach program improves the return on the randomized outreach program by 74%–96% or $1.6 million to $2 million. Thus, outreach marketing provides a substantial positive payoff to the health care system.

Keywords

causal forests, machine learning, personalized health care marketing, cancer screening, randomized field experiment

Online supplement: https://doi.org/10.1177/0022242920913025

In 2018, over 1.7 million new cases of cancer were diagnosed in the United States, and the cost of cancer care surpassed $147 billion (National Cancer Institute 2018). Following the guidelines of the National Comprehensive Cancer Network (2018), health care institutions encourage at-risk patients to undergo regular screening, as this opens the door for early detection, more cost-effective treatment options, and better recovery prognosis. Regular screening reduces mortality rates for lung (20% drop; Humphrey et al. 2013), breast (20%–40% drop; Oeffinger et al. 2015), and liver (37% drop; Zhang, Yang, and Tang 2004) cancers. Moreover, cancer screening can reduce annual treatment costs by nearly $5,000 (Benoit, Grönberg, and Naslund 2001).

Health care institutions invest heavily in direct-to-patient outreach interventions to increase screening completion among at-risk patients. For example, Johns Hopkins Hospital's cancer center uses emails, letters, seminars, and community events to encourage screening completion among patients (Johns Hopkins Medicine 2019). With 1.7 million outreach interventions launched in 2015, and $123 million spent on prevention and

Yixing Chen is Assistant Professor of Marketing, Mendoza College of Business, University of Notre Dame, USA (email: ychen43@nd.edu). Ju-Yeon Lee is Assistant Professor of Marketing, Ivy College of Business, Iowa State University, USA (email: leejy@iastate.edu). Shrihari (Hari) Sridhar is Professor of Marketing, Joe Foster '56 Chair in Business Leadership, Mays Business School, Texas A&M University, USA (email: ssridhar@mays.tamu .edu). Vikas Mittal is J. Hugh Liedtke Professor of Marketing, Jones Graduate School of Business, Rice University, USA (email: vmittal@rice.edu). Katharine McCallister is Population Science Program Manager, University of Texas Southwestern Medical Center, USA (email: katharine.mccallister@ utsouthwestern.edu). Amit G. Singal is Associate Professor of Internal Medicine, Clinical Chief of Hepatology, University of Texas Southwestern Medical Center, USA (email: amit.singal@utsouthwestern.edu). Amit G. Singal has served as a consultant for Wako Diagnostics, Roche, Exact Science, and Glycotest.

education efforts,[1] only 8% of U.S. adults over 35 years old utilize preventive services (Borsky et al. 2018). This percentage is too low. Health care institutions face three challenges in improving outreach effectiveness.

First, most studies examine only the main effects of medical interventions (e.g., Singal, Pillai, and Tiro 2014; Singal et al. 2019), neglecting variation due to patient demographics, health status, visit history, health system accessibility, and neighborhood socioeconomic status (Figueroa and Jha 2018). By incorporating this heterogeneity in patient response to outreach interventions, health care institutions can implement "personalized health care marketing" and boost outreach effectiveness. Second, medical scholars typically compare the relative efficacy of outreach interventions using a single-period research design (e.g., Basch et al. 2006). Given the importance of regular screening compliance over multiple periods (Chubak and Hubbard 2016), it is critical to evaluate screening compliance over multiple periods. Third, quantifying the return on outreach interventions to incorporate the health benefits and financial cost of interventions will help health care institutions communicate the tangible value they bring to the community and enable funding agencies to sustain these interventions (Andersson et al. 2008). As the director of cancer education for the Stanford Cancer Center notes, "durable, long-term solutions will require a substantial investment in academic/community partnerships to improve cancer education" (Conger 2010).

To addresses these three challenges, we use a multiperiod randomized field experiment conducted at a large hospital system with at-risk patients for hepatocellular carcinoma (HCC), the most common type of primary liver cancer (Singal et al. 2019). Patients were randomly assigned (1:1:1) to three conditions: usual care, outreach alone, or outreach with patient navigation. Usual care is the baseline condition in which physicians offer preventive care recommendations at their discretion during a patient's usual care visits. As we describe subsequently, outreach alone and outreach with patient navigation provide two different levels of marketing using outreach mails, outreach calls, and customized motivational education by trained patient navigators. The focal outcome is the patient's screening completion status within 6 months (Period 1), 6–12 months (Period 2), and 12–18 months (Period 3) of the initial randomization. This enables an investigation of the impact of outreach interventions on regular screening compliance. We evaluated screening completion status every 6 months, as this interval has been demonstrated to increase early detection and survival compared with longer screening intervals (Santi et al. 2010). To incorporate patient heterogeneity, we iteratively construct the focal covariates based on the extant medical literature and pragmatic considerations that the study design affords, including patients' demographics, health status, visit history,

health system accessibility, neighborhood socioeconomic status, and prior screening compliance.

Relative to the baseline condition, outreach alone (outreach with patient navigation) increases screening completion rates by 10–20 (13–24) percentage points, but the effectiveness of the two outreach interventions does not significantly differ. Central to this article, the similarity in these main effects masks considerable heterogeneity in outreach effectiveness due to patient-level differences (Blanchard et al. 2012). We uncover patient-level treatment effects of these two interventions using causal forests, a state-of-the-art development in the machine learning and economics literature (Wager and Athey 2018). Results show the following: (1) compared with outreach alone, outreach with patient navigation induces a higher proportion of patients with significant positive heterogeneous treatment effects in Periods 2 (9%) and 3 (23%); (2) the increased screening completion from outreach alone or outreach with patient navigation is higher for patients who are female, are part of a racial/ethnic minority, have a better health status, have a more frequent visit history, are covered by medical-assistance insurance, reside in closer proximity to clinics, and reside in a more populated neighborhood; (3) the increase in screening completion due to outreach alone is higher for patients who are younger, commute faster, and reside in a neighborhood with more public insurance coverage; in contrast, the increase in screening completion as a result of outreach with patient navigation is higher for patients who are older and reside in a higher-income neighborhood.

Incorporating patient-level differences in their responsiveness to outreach interventions, and a well-established scheme of cost–benefit calculation that quantifies health benefits and financial costs associated with outreach interventions (e.g., Goossens et al. 2017), we assign patients to the baseline, outreach-alone, or outreach-with-patient-navigation condition in each period on the basis of their predicted treatment effect and predicted net return. As a result, the commensurate return on the patient-level targeted outreach program is $3,704,270–$4,167,419 when extrapolated to 3,217 eligible patients in the hospital's database. The targeted outreach program improves the return on the randomized outreach program ($2,130,921) by 74%–96%.

We make several contributions to marketing theory and health care practice. First, the literature on marketing interventions in health care (Table 1, Panel A) typically relies on experimentally manipulated moderators, such as test accuracy (Luce and Kahn 1999) or consumer goals (Wang, Keh, and Bolton 2010); while theoretically interesting, they are impractical for health care institutions to implement. Health care institutions can readily utilize observable patient characteristics— such as ethnicity, visit history, and insurance coverage—that are of theoretical relevance. The bulk of the marketing literature has focused on attitudinal consequences using self-reports of behavioral intentions (e.g., Bolton et al. 2008), risk perceptions (e.g., Menon, Block, and Ramanathan 2002), and attitudes (e.g., Basil and Brown 1997) in a lab setting. While insightful, they are of little practical relevance to addressing

---

[1] See https://www.cancer.org/content/dam/cancer-org/online-documents/en/pdf/infographics/where-does-your-money-go-infographic-print.pdf (accessed February 25, 2020).

Table 1. Literature Review on Cancer Care and Preventive Care.

		Dependent Variable	Independent Variable	Approaches to Understand Heterogeneity of Treatment Effect			
Reference	Context (Sample)	Objective Behavior (vs. Self-Reported)	Degree of Intervention	Use of Patient Characteristics as Sources of Heterogeneity	Evaluation of Patient-Level Treatment Effectiveness	Identification Method	Evaluation of Return on Intervention
Our study	Liver cancer (patients with suspected or documented risks of cirrhosis)	✓ Actual behavior: screening completion	Three types of outreach (no outreach, outreach alone, outreach with patient navigation)	✓ Demographics, health status, visit history, health system accessibility, neighborhood socioeconomic status	✓	Randomized field experiment	✓
A: Marketing Literature							
Block and Keller (1995)	Sexually transmitted disease and skin cancer (college students)	X Self-reported perceptions: in-depth processing, attitude, intentions	Two types of efficacy (high vs. low)	X	X	Lab experiment	X
Keller and Block (1996)	Smoking (college students)	X Self-reported perceptions: persuasiveness	Two types of messages (low vs. high fear evoking)	X	X	Lab experiment	X
Jayanti and Burns (1998)	Primary care (patients)	X Self-reported intentions: preventive health care behavioral intention	No intervention (health knowledge/motivation, health consciousness, self-efficacy/response efficacy, health value)	X	X	Survey	X
Luce and Kahn (1999)	Breast cancer (women in mammography waiting rooms)	X Self-reported perceptions: perceptions of vulnerability and test inaccuracy	Two types of results (normal vs. false alarm)	X	X	Lab experiment	X
Cox and Cox (2001)	Breast cancer (women from social volunteer organizations)	X Self-reported perceptions: attitudes and beliefs toward the target behavior	Two types of messages (statistical vs. anecdotal)	X	X	Lab experiment and in-depth interviews	X
Keller, Lipkus, and Rimer (2002)	Breast cancer (older women and college students)	X Self-reported perceptions: risk perception	Two types of mood (depressives vs. nondepressives)	X	X	Lab experiment	X
Menon, Block, and Ramanathan (2002)	Hepatitis C (college students)	X Self-reported perceptions: intention, concern, self-positivity bias, self-risk estimates, message effectiveness	Two types of message cues (frequent vs. infrequent)	X	X	Lab experiment	X
Kahn and Luce (2003)	Breast cancer (women in mammography waiting rooms)	X Self-reported perceptions: stress, disutilities of harm, retest intention	Two types of results (normal vs. false alarm)	X	X	Lab experiment	X

Table 1. (continued)

Reference	Context (Sample)	Dependent Variable Objective Behavior (vs. Self-Reported)	Independent Variable Degree of Intervention	Approaches to Understand Heterogeneity of Treatment Effect Use of Patient Characteristics as Sources of Heterogeneity	Evaluation of Patient-Level Treatment Effectiveness	Identification Method	Evaluation of Return on Intervention
Bowman, Heilman, and Seetharaman (2004)	Prescription drugs for chronic ailments (patients)	✓ Approximated behavior: approximated drug consumption	No intervention (advertising cues, salience/mindfulness, benefits/costs of compliance, threats)	X	X	Secondary data	X
Bolton, Cohen, and Bloom (2006)	Smoking cessation aids (college students)	X Self-reported perceptions: risk perception, smoking-cessation intention	Two types of message (remedy vs. none)	X	X	Lab experiment and field study	X
Bolton et al. (2008)	High cholesterol (older men) and weight management (staff and students)	X Self-reported perceptions: healthy lifestyle intentions, risk/health perceptions	Two types of health remedy (drug vs. supplements)	X	X	Mailed surveys and lab experiments	X
Du, Sen, and Bhattacharya (2008)	Oral health (disadvantaged Hispanic families)	X Self-reported perceptions: belief, behaviors	Two types of health education (participation vs. no participation)	✓ Acculturation (self-reported)	X	Focus group interviews and quasiexperiment	X
Cox, Cox, and Mantel (2010)	Prescription drug for the early detection of skin cancer (young adults)	X Self-reported perceptions: behavioral intention, attitude, perception	Two (high vs. low frequency) × two (high vs. low severity) types of side effects	X	X	Lab experiment	X
Wang, Keh, and Bolton (2010)	Western versus Eastern medicine (college students)	X Self-reported perceptions: preference for health remedies, healthy lifestyle intentions	Two (high vs. low) diagnosis uncertainty × two (alleviating symptoms vs. curing illness) goals	X	X	Lab experiment and field study	X
Samper and Schwartz (2012)	Skin cancer and flu shots (adults)	X Self-reported perceptions: risk perception, intentions to consume care	Two levels of medication price (high vs. low)	X	X	Lab experiment	X
Lisjak and Lee (2014)	Unprotected sex, kidney diseases and chlamydia tests (college students)	X Self-reported perceptions: intention, perceived vulnerability	Two levels of depletion (depletion, nondepletion)	X	X	Lab experiment	X
B: Medical and Health Care Literature							
Schut et al. (1997)	Grief counseling (widows and widowers)	X Self-reported perception: psychological distress	Three types of intervention (control, emotion-focused, problem-focused grief counseling)	✓ Gender	X	Randomized field experiment	X

Table I. (continued)

Reference	Context (Sample)	Dependent Variable Objective Behavior (vs. Self-Reported)	Independent Variable Degree of Intervention	Approaches to Understand Heterogeneity of Treatment Effect		Identification Method	Evaluation of Return on Intervention
				Use of Patient Characteristics as Sources of Heterogeneity	Evaluation of Patient-Level Treatment Effectiveness		
Murray et al. (2004)	Drug use (elderly patients with chronic heart failure)	✓ Objective behavior: medication adherence	Two types of intervention (usual care, pharmacist support)	X	X	Randomized field experiment	X
Unger et al. (2004)	Smoking prevention (middle school students)	X Self-reported behavior: lifetime smoking	Two smoking prevention curricula (standard, multicultural curriculum)	✓ Ethnicity, gender	X	Randomized field experiment	X
Sullivan et al. (2005)	Asthma care (children with asthma)	X Self-reported estimated behavior: symptom-free days	Three types of intervention (usual care, peer leader-based physician behavior change intervention, planned asthma care intervention)	X	X	Randomized field experiment	✓
Basch et al. (2006)	Colorectal cancer (African American patients)	✓ Objective behavior: screening completion	Two types of intervention (tailored telephone outreach, mail outreach)	X	X	Randomized field experiment	X
Ziegelmann, Lippke, and Schwarzer (2006)	Orthopedic rehabilitation (rehabilitation center patients)	X Self-reported intentions: coping/action plans, physical activities	Two types of intervention (standard care, interviewer-assisted planning)	✓ Age	X	Randomized field experiment	X
Wang et al. (2007)	Depression (employees)	X Self-reported perception: depression severity	Two types of outreach (control, telephonic outreach)	✓ Health conditions	X	Randomized field experiment	X
Andersson et al. (2008)	Liver disease (patients with compensated cirrhosis)	✓ Simulated behavior: screening completion	No intervention (intervention is not the main focus)	X	X	Simulation	✓
Ferrante, Chen, and Kim (2008)[a]	Breast cancer (women with suspicious mammogram results)	✓ Objective behavior: diagnostic interval	Two types of outreach (no outreach, outreach with patient navigation)	✓ Ethnicity, health conditions	X	Randomized field experiment	X
Gabram et al. (2008)	Breast cancer (African American women)	✓ Objective behavior: screening completion	One condition (community education/outreach initiatives)	X	X	Cross-sectional study	X
Muller et al. (2009)	Colorectal cancer (adults aged 50 to 80 years)	✓ Objective behavior: screening completion	Three types of outreach (no outreach, letter reminder, email reminder)	X	X	Randomized field experiment	X
Humiston et al. (2011)	Influenza immunization (older patients)	✓ Objective behavior: immunization completion	Two types of outreach (standard of care, outreach intervention)	✓ Ethnicity	X	Randomized field experiment	X

(continued)

5

Table 1. (continued)

Reference	Context (Sample)	Dependent Variable: Objective Behavior (vs. Self-Reported)	Independent Variable: Degree of Intervention	Approaches to Understand Heterogeneity of Treatment Effect: Use of Patient Characteristics as Sources of Heterogeneity	Evaluation of Patient-Level Treatment Effectiveness	Identification Method	Evaluation of Return on Intervention
Gupta et al. (2013)	Colorectal cancer (uninsured patients)	✓ Objective behavior: screening completion	Three types of outreach (no outreach, mailed invitation to return a test, mailed invitation to schedule colonoscopy)	X	X	Randomized field experiment	X
Wigg et al. (2013)	Liver disease (patients with cirrhosis and complications from chronic liver failure)	✓ Objective behavior: outpatient care attendance rates	Two types of intervention (absence or presence of chronic disease management programs)	X	X	Randomized field experiment	X
Wetherell et al. (2013)	Anxiety disorders (patients diagnosed with anxiety disorder)	X Self-reported perception: anxiety level	Two types of intervention (usual care, computer-assisted collaborative care)	✓ Age	X	Randomized field experiment	X
Baker et al. (2014)	Colorectal cancer (mostly Hispanic female patients)	✓ Objective behavior: screening completion	Two types of outreach (no outreach, outreach with patient navigation)	X	X	Randomized field experiment	X
Beste et al. (2015)	Liver cancer (patients with cirrhosis who have visited Veterans Affairs facilities)	✓ Objective behavior: screening completion	Two types of intervention (no reminder, point-of-care reminder)	X	X	Quasiexperiment	X
Goldman et al. (2015)	Colorectal cancer (patients with no documentation of screening)	✓ Objective behavior: screening completion	Two types of outreach (usual care, multifaceted intervention)	✓ Visit history, time	X	Randomized field experiment	X
Singal et al. (2017)	Liver cancer (patients with suspected or documented risks of cirrhosis)	✓ Objective behavior: screening completion	Three types of outreach (no outreach, outreach alone, outreach with patient navigation)	X	X	Randomized field experiment	X
Coronado et al. (2018)	Colorectal cancer (adults aged 50 to 74 years)	✓ Objective behavior: screening completion	Two types of outreach (no outreach, mailed outreach)	X	X	Randomized field experiment	X
McCarthy et al. (2018)	Contraception (young people in Tajikistan)	✓ Objective behavior: use of contraception	Three types of intervention (control, mobile app messaging intervention)	✓ Gender	X	Randomized field experiment	X
Singal et al. (2019)	Liver cancer (patients with suspected or documented risks of cirrhosis)	✓ Objective behavior: screening completion	Three types of outreach (no outreach, outreach alone, outreach with patient navigation)	✓ Race, visit history, health status	X	Randomized field experiment	X

[a]The authors note that they were unable to find the heterogeneity of treatment effect due to the small sample size in each subgroup.

Notes: References are provided in Web Appendix H.

actual health care behaviors such as screening completion among real patients.

Second, we extend the medical literature on cancer outreach effectiveness, which has focused primarily on the main effects of cancer outreach interventions from randomized field studies (Table 1, Panel B). The causal forests approach provides a practical way to improve the efficacy and external validity of field experiments by systematically exploring the treatment-effect heterogeneity across intervention types, across patient subgroups, and over time without prespecifying the sources of heterogeneity (Cook and Campbell 1979; Lynch 1982).

Third, we provide insights into what patient subgroup benefits more (less) from outreach interventions, offer ways to customize the interventions, and help practitioners allocate limited financial resources to those with the largest potential gains. For example, while outreach programs typically target diverse, socioeconomically difficult-to-reach disadvantaged patient populations to improve their health outcomes (Singal et al. 2019), patients more responsive to outreach interventions tend to be female, be part of a minority, be in good health status, have more frequent visit history, be covered by medical-assistance insurance, reside in closer proximity to clinics, and reside in more populated neighborhoods. Thus, simply targeting one or two patient characteristics may not maximize the gains from the outreach interventions.

Fourth, our approach provides a roadmap for implementing personalized health care marketing by customizing outreach interventions and quantifying the return on such interventions. Using patient-level treatment effect estimates with valid confidence intervals, we not only provide a tool that can recommend the most suitable intervention for each patient given their profile but also provide an individual-level cost–benefit analysis to measure the return on personalized health care marketing investments.

Institutional Setting, Study Design, and Data

Institutional Setting: Cancer Outreach and the Importance of Regular Screening

Our field experiment is based on the cancer outreach efforts of a large hospital system to increase regular screening completion for early detection of HCC, the most common form of liver cancer, among patients with higher risk of HCC. Most patients with liver cancer do not display symptoms until it reaches an advanced stage; they often miss the time window during which treatment options, such as transplant and surgical resection, are effective. The five-year survival rate for early-stage liver cancer patients who undergo surgery is 60%–70%, while the five-year relative survival rate for liver cancer is 18% (American Cancer Society 2019). Yet, the utilization rate of HCC screening is below 20% in the general cirrhotic population, and even lower among low-socioeconomic-status and non-Caucasian patients (Singal et al. 2012).

The outreach program was designed to promote regular screening (i.e., obtain a screening every six months). The

six-month screening interval is in line with the evidence-based recommendations issued by the American Association for the Study of Liver Diseases and National Comprehensive Cancer Network (e.g., Marrero et al. 2018; Tzartzeva et al. 2018). It was initially based on tumor doubling times and (1) is better for early detection than longer intervals (e.g., 12 months; Santi et al. 2010) but worse than shorter intervals (e.g., 3 months; Trinchet et al. 2011) and (2) minimizes patient and provider burden (Bruix and Sherman 2011).

The hospital system conducted a randomized trial between December 2014 and March 2017. The study was approved by the University of Texas Southwestern Medical Center Institutional Review Board. The trial protocol is available on clinical trials.gov (NCT02312817), where the study is registered. The random assignment (1:1:1) consisted of one baseline condition (no outreach) and two conditions with outreach interventions (outreach alone and outreach with patient navigation), with the outcome being HCC screening completion status.[2]

Study Design

Sample. The eligibility criteria for patient inclusion using established norms have been developed in the medical field (for details, see Web Appendix A1). From the 3,217 eligible patients in the hospital's database, 1,800 patients were randomly selected for the study.[3]

Focal independent variable: intervention type. As we summarize in Figure 1, each patient was randomly assigned to one of three conditions in a 1:1:1 ratio:

- **No outreach or usual care (baseline condition)**: Patients received visit-based HCC screening as recommended by primary or specialty care providers and were not contacted by the outreach marketing team. For patients who scheduled ultrasounds, the hospital system placed automated reminder telephone calls two days before the ultrasound appointments.
- **Outreach-alone intervention**: As in the baseline condition, patients were eligible for usual care, as offered through their usual outpatient encounters. Patients were also mailed a one-page letter, which contained information on the risk of HCC in patients with cirrhosis and the benefits and risks of HCC screening, a brief summary of the screening procedure, and a recommendation to the patient to make an appointment for an ultrasound (for details, see Web Appendix A2). To increase participation, the staff then made outreach calls to nonresponders (i.e., patients with returned mail and those who did not respond to mailed invitations within two to four weeks).

---

[2] The hospital system is the sole safety-net provider for Dallas County, which minimizes omitted variable bias that could emanate from competitive efforts by other organizations in the area.

[3] The hospital system obtained a waiver of informed consent to minimize volunteer bias.

During telephone calls, trained research staff followed standardized scripts. Mails and telephone calls were in English or Spanish, depending on patients' preferences. In addition, the hospital system placed automated reminder telephone calls two days before appointments for patients who scheduled ultrasounds.

- **Outreach-with-patient-navigation intervention:** Like patients in the baseline condition and those in the outreach-alone condition, patients in this condition were eligible for care as offered through their usual outpatient encounters. Patients in this condition had an experience identical to those in the outreach-alone condition, with two additions: (1) a telephone script used during outreach telephone calls and (2) an additional reminder call from the research staff. During telephone calls, if patients in this condition declined to make an appointment for screening, the research staff used a standardized telephone script to identify potential barriers and then provided customized motivational messages to encourage screening participation. Examples of barriers include preparation involved, pain during the test, and so on (for details, see Web Appendix A3). For instance, if a patient is concerned about the preparation required for the screening, the research staff alleviates this concern by stating, "A liver ultrasound is a quick procedure. The ultrasound usually takes less than 30 minutes and the appointment should take around one hour from start to finish." For scheduled ultrasounds, the hospital system's research staff called the patients five to seven days before the appointments to provide a reminder, address any concerns, and reschedule the appointment if needed. For these patients, the hospital system also placed automated reminder telephone calls two days before the ultrasound appointments. Overall, as shown in Figure 1, this condition is the most intense and comprehensive intervention in the study.

Multiperiod study design and sample sizes in periods 2 and 3. To encourage regular screening completion, the study repeated the outreach-alone and outreach-with-patient-navigation interventions during each of the three periods. We define Period 1 as the time within 6 months of the first randomization, Period 2 as the time between month 6 and month 12 since the first randomization, and Period 3 as the time between month 12 and month 18 since the first randomization. In summary, the hospital system undertook the outreach interventions in all three periods, each period being six months apart, and each patient belonging to the same condition across the three periods. The goal of this design is to encourage screening in each period.

Once a patient has completed the screening in the first period, the patient does not exit the pool and is contacted in the second and third six-month periods. There are two exceptions to the repeated interventions: (1) if the patient completes the screening and is diagnosed with HCC during the experiment, the patient exits the pool as the providers must refer the patient for HCC treatment instead of routine

screening; (2) if the patient completes the screening and dies during the course of the experiment, the patient cannot complete the screening in later periods. As a result, the sample size is 1,800 for Period 1, 1,772 for Period 2, and 1,743 for Period 3.[4] The sample sizes in the baseline, outreach-alone, and outreach-with-patient-navigation conditions are (600, 600, 600) for Period 1, (591, 592, 589) for Period 2, and (577, 584, 582) for Period 3.

Dependent variable: screening completion status. Screening completion status is measured as a patient getting an abdominal imaging screening test (1) or not (0).[5] We observe the dependent variable for each patient in Periods 1, 2, and 3.

Constructing Focal Covariates: An Iterative Approach

Taking theoretical and pragmatic considerations into account, we followed a four-step iterative approach to determine the focal covariates that inform patient heterogeneity in response to outreach interventions. This process of including covariates starts from original yet tentative variables available to researchers and is informed by a multifaceted understanding of theory models, prior studies, research questions, and practice. The approach resembles a theory-in-use process (e.g., Zeithaml et al. 2020) that iteratively intertwines exploratory and confirmatory research to incorporate the interplay of heterogeneity with treatment.

- *Step 1: Utilize original variables.* We begin with the variables that are available in the electronic medical record system (EMR) and are relevant to practitioners and well-documented in medical research (and thus relevant to academic scholars).[6] These systems store and track key patients' information such as patient demographics (e.g., Wetherell et al. 2013), health status (e.g., Ferrante, Chen, and Kim 2008), and visit history record (Skolnik 2011). As Table 1 shows, previous studies in health care have analyzed these "ready-for-use" variables in the EMR (e.g., Humiston et al. 2011; McCarthy et al. 2018).
- *Step 2: Construct theoretically relevant variables.* We use the information available in the EMR to construct new variables that are not captured by the raw unrefined data but draw on theories such as health belief model and protection motivation theory (e.g., Lisjak and Lee 2014; Moorman and Matulich 1993). Thus, a patient's health insurance and location information proxy their

---

[4] Out of 28 patients excluded in Period 2, 12 were excluded because they were diagnosed with HCC in Period 1, and 16 were deceased. Out of 57 patients excluded in Period 3, 23 were excluded due to being diagnosed with HCC in Period 2, and 38 were deceased (4 of them were both diagnosed with HCC and deceased).

[5] The abdominal imaging screening includes an ultrasound, magnetic resonance imaging (MRI), or computed tomography (CT).

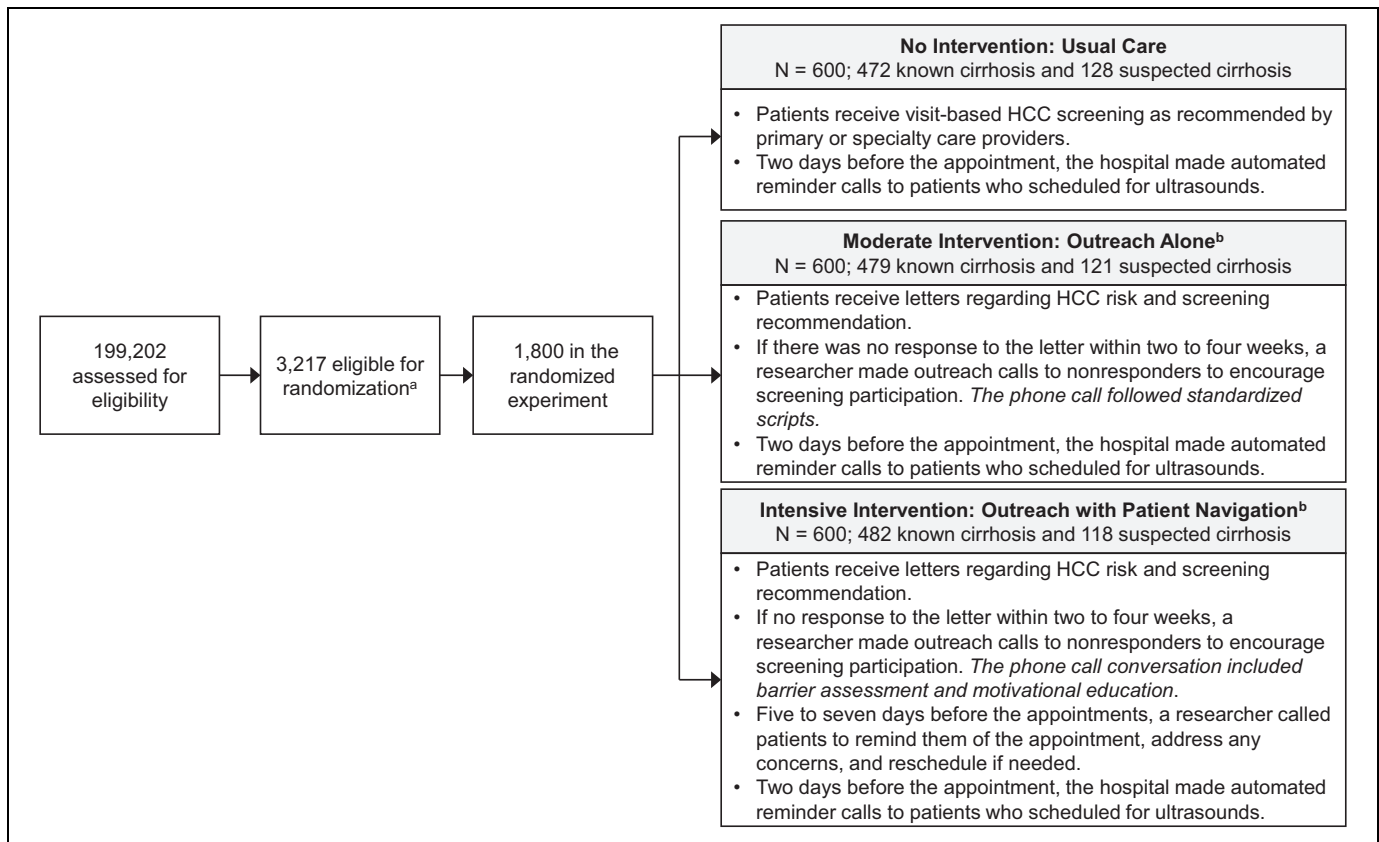[6] In 2017, 85.9% of office-based physicians in the United States used an EMR system (https://www.cdc.gov/nchs/fastats/electronic-medical-records.htm).

No Intervention: Usual Care
N = 600; 472 known cirrhosis and 128 suspected cirrhosis

- Patients receive visit-based HCC screening as recommended by primary or specialty care providers.
- Two days before the appointment, the hospital made automated reminder calls to patients who scheduled for ultrasounds.

Moderate Intervention: Outreach Alone[b]
N = 600; 479 known cirrhosis and 121 suspected cirrhosis

- Patients receive letters regarding HCC risk and screening recommendation.
- If there was no response to the letter within two to four weeks, a researcher made outreach calls to nonresponders to encourage screening participation. *The phone call followed standardized scripts.*
- Two days before the appointment, the hospital made automated reminder calls to patients who scheduled for ultrasounds.

Intensive Intervention: Outreach with Patient Navigation[b]
N = 600; 482 known cirrhosis and 118 suspected cirrhosis

- Patients receive letters regarding HCC risk and screening recommendation.
- If no response to the letter within two to four weeks, a researcher made outreach calls to nonresponders to encourage screening participation. *The phone call conversation included barrier assessment and motivational education.*
- Five to seven days before the appointments, a researcher called patients to remind them of the appointment, address any concerns, and reschedule if needed.
- Two days before the appointment, the hospital made automated reminder calls to patients who scheduled for ultrasounds.

199,202 assessed for eligibility → 3,217 eligible for randomization[a] → 1,800 in the randomized experiment

Figure 1. Study design: Cancer outreach interventions.
[a]The following patients were excluded: 185,539 patients who did not meet cirrhosis criterion, 9,921 patients with comorbid conditions or with Child C cirrhosis, 405 patients with the history of HCC or suspicious mass on imaging, 78 patients whose language is not English or Spanish, and 42 patients with no contact information.
[b]Patients in the outreach-alone and outreach-with-patient-navigation conditions could receive usual care.

insurance coverage (financial access to care) and proximity to clinics (geographical access to care). This is consistent with the research showing that health system accessibility and "improving health system accessibility across the socio-economic spectrum" (Australian Digital Health Agency 2018, p. 19) is a strategic priority for policy makers (Provincial Health Services Authority 2011).

- *Step 3: Explore external secondary data sources.* To supplement the previous steps, we also gather additional data from external secondary sources. Socioeconomic factors can help marketing researchers develop a better understanding of understudied and underserved consumers (MacInnis et al. 2020). We collect data on each patient's neighborhood socioeconomic status—including educational attainment, income, commute time, private/public health insurance coverage, employment status, and population—by collecting zip-code-level data from American Community Survey.

- *Step 4: Incorporate contextually relevant variables.* Along with variables that are static in nature, we include each patient's screening compliance in prior periods. Incorporating cancer screening compliance across

different periods (Chubak and Hubbard 2016) captures the temporal variation in screening completion. It also informs us how outreach effectiveness might vary due to patients' prior behavioral pattern.

In summary, we include six sets of patient characteristics: (1) demographics including age, gender (coded as 1 if a patient is female, 0 otherwise), ethnicity (non-Hispanic Caucasian, Hispanic, Non-Hispanic African American, or other/unknown), and primary language (English, Spanish, or other); (2) health status, which includes Child-Pugh B (coded as 1 if Child-Pugh score is higher than 6, 0 otherwise), Charlson Comorbidity Index, presence of documented cirrhosis (coded as 1 if yes), etiology of liver disease (hepatitis C, hepatitis B, alcohol, nonalcoholic steatohepatitis, or other); (3) visit history, which includes the number of primary care visits in the year prior to cohort entry and receipt of hepatology care (coded as 1 if the patient received the hepatology care prior to cohort entry, 0 otherwise); (4) health system accessibility, which includes insurance coverage (commercial, Medicaid, medical assistance/charity, Medicare, self-pay, or unknown) and proximity to clinics (coded as 1 if there are more than three clinics in the zip code that matches the first three digits

of the zip code[7] where the patient resides, 0 otherwise); (5) neighborhood socioeconomic status, which includes educational attainment (percentage with a bachelor's degree or higher), income (per capita income), average commute time, insurance coverage (percentage with a private or public health insurance plan), unemployment rate, and population measured at the three-digit zip code level[8]; and (6) screening completion status in the prior period(s) (coded as 1 if a patient completes the screening test in Periods 1 or 2, 0 otherwise). Table 2 describes each variable, its operationalization, and descriptive statistics. Web Appendix B compares the means of all variables across three conditions. Differences are statistically nonsignificant, showing that random assignment was successful.

Model-Free Evidence

Figure 2, Panel A, shows the number of patients who completed screening in different periods. Whereas 435 patients (24%) completed the screening only once during the three periods, 660 patients (37%) did so more than once. For all three periods, more patients completed the screening with outreach alone (102) and outreach with patient navigation (134) than the usual care (36) condition. The evidence suggests that both interventions increase HCC screening.

Figure 2, Panel B, shows screening completion rates in each condition in each period after excluding the patients who were deceased or diagnosed with HCC in the previous period(s). In Period 1, 25% in the no-outreach condition, 45% in the outreach-alone condition, and 48% in the outreach-with-patient-navigation condition underwent screening. The screening completion rate in the outreach-alone condition (difference = .198, $p < .01$) and the outreach-with-patient-navigation condition (difference = .232, $p < .01$) is significantly higher than the no-outreach condition. Results in Periods 2 and 3 show a similar pattern. Comparing the screening completion rate in the outreach-alone condition and the outreach-with-patient-navigation condition, there is no statistically significant difference in Period 1 (difference = .033, n.s.) or Period 2 (difference = .033, n.s.), and Period 3 (difference = .051, $p < .10$). The model-free evidence suggests that both outreach conditions outperform the baseline condition but do not differ in effectiveness relative to each other.[9]

Empirical Strategy

Causal Forest Estimation of Patient-Level Treatment Effects

To draw inferences about the causal effect of different interventions, researchers typically estimate and compare the average treatment effects (i.e., main effects) of randomized interventions. Such a comparison may not consider that treatment effects vary across subgroups within and across treatment conditions. Moreover, to avoid searching for particularly responsive subgroups, medical researchers must register preanalysis protocols for clinical trials to specify which subgroups will be analyzed. Such protocols may fail to identify strong but unexpected treatment-effect heterogeneity, especially in emergent fields in which moderators are ex ante ambiguous. We use causal forests to address these two challenges (Wager and Athey 2018). Causal forests enable nonparametric estimation of patient-level treatment effects with valid asymptotic confidence intervals, without restrictions on the number of covariates or the need for a larger number of experimental conditions or repeated measures. Causal forests also alleviate concerns regarding spurious treatment-effect heterogeneity due to searching for particularly responsive subgroups (Web Appendix D1 and D2 compare causal forests with several established approaches). Next, we outline the potential outcome framework, followed by an overview of causal forests.

Potential outcome framework. For illustration purposes, we consider the case of one period and the outreach-alone intervention (treatment condition) compared with no outreach/usual care (control condition). For a set of independent and identically distributed patients $i = 1, \ldots, n$, we observe the outcome of interest Y_i (screening completion), treatment assignment W_i (i.e., whether the patient is assigned to the outreach-alone or no-outreach condition), and vector of patient characteristics X_i (e.g., patient demographics, health status). Following the potential outcome framework (Rubin 1974), for each patient i, there are two potential outcomes: if a patient is assigned to the treatment condition, we observe the outcome $Y_i = Y_{i1}$, and if the patient is assigned to the control condition, we observe $Y_i = Y_{i0}$. We define the conditional average treatment effect (CATE) (i.e., treatment effect at x) to assess whether the treatment effect is heterogeneous among subgroups:

$$\tau(x) = E[Y_{i1} - Y_{i0} | X_i = x]. \tag{1}$$

The fundamental challenge to identifying the CATE is that we only observe one of the two potential outcomes: Y_{i1} and Y_{i0}. Thus, we must invoke the assumption of unconfoundedness to estimate the CATE (Rosenbaum and Rubin 1983). As patients are

---

[7] According to Health Insurance Portability and Accountability Act, we are not allowed to obtain patients' identifiable location information such as address and zip code. Thus, we obtain the deidentified version (i.e., the first three digits of the zip code).

[8] Because we observe only the first three digits of patients' zip code, all zip-code-level covariates are aggregated to the three-digit level by calculating the sum (i.e., population) or mean (percentage with a bachelor's degree or higher, mean travel time to work, and per capita income) across all five-digit zip codes that share the same first three digits.

[9] Web Appendix C suggests that, on average, the significant difference in "no-show rates" in Period 3 seems to drive the main effects of outreach interventions on screening completion, while, statistically, there is no difference in scheduling rates across two outreach interventions across all

---

three periods. This finding also seems to suggest that reminder calls made by the research staff in the outreach-with-patient-navigation condition may guide patients toward screening completion by lowering the probability that they would not show up. However, pinning down the exact mechanisms is beyond the scope of this research.

Table 2. Summary Statistics.

Variable	Definition	Mean	SD	Min	Max
Dependent Variable					
Completion in Period 1	Whether a patient underwent an abdominal imaging screening test, which includes ultrasound, MRI, and CT, in Period 1 (0–6 months after cohort entry). Coded as 1 if the patient completed, and 0 otherwise.	39.3%	—	0	1
Completion in Period 2	Whether a patient underwent an abdominal imaging screening test, which includes ultrasound, MRI, and CT, in Period 2 (6 months and 1 day–12 months after cohort entry). Coded as 1 if the patient completed, and 0 otherwise.	38.4%	—	0	1
Completion in Period 3	Whether a patient underwent an abdominal imaging screening test, which includes ultrasound, MRI, and CT, in Period 3 (12 months and 1 day–18 months after cohort entry). Coded as 1 if the patient completed, and 0 otherwise.	34.9%	—	0	1
Independent Variable					
Outreach intervention	A baseline condition and two outreach intervention types:				
	1. No outreach (usual care)	33.3%	—	0	1
	2. Moderate outreach (outreach alone)	33.3%	—	0	1
	3. Intensive outreach (outreach with patient navigation)	33.3%	—	0	1
Demographics					
Age (years)	Age of the patient at cohort entry	55.3	10.5	21	90
Gender	Gender of the patient (0 = male, 1 = female)	40.6%	—	0	1
Ethnicity					
Non-Hispanic Caucasian	Non-Hispanic Caucasian = 1, otherwise = 0	28.3%	—	0	1
Hispanic	Hispanic = 1, otherwise = 0	37.8%	—	0	1
Non-Hispanic African American	Non-Hispanic African American = 1, otherwise = 0	32.1%	—	0	1
Other/unknown	Other/unknown = 1, otherwise = 0	1.7%	—	0	1
Language					
English	English = 1, otherwise = 0	76.9%	—	0	1
Spanish	Spanish = 1, otherwise = 0	22.7%	—	0	1
Other	Other = 1, otherwise = 0	.3%	—	0	1
Health Status					
Child Pugh B	Whether a patient is Child Pugh B, coded as 1 if Child Pugh Score > 6, 0 otherwise.	28.3%	—	0	1
Charlson Comorbidity Index	Charlson Comorbidity Index Score	2.9	2.4	0	12
Documented cirrhosis	Cirrhosis diagnosis (0 = suspected cirrhosis, 1 = known cirrhosis)	79.6%	—	0	1
Etiology of liver disease					
Hepatitis C	Hepatitis C = 1, otherwise = 0	51.0%	—	0	1
Hepatitis B	Hepatitis B = 1, otherwise = 0	3.4%	—	0	1
Alcohol	Alcohol = 1, otherwise = 0	17.6%	—	0	1
Nonalcoholic steatohepatitis	Nonalcoholic steatohepatitis = 1, otherwise = 0	16.6%	—	0	1
Other	Other = 1, otherwise = 0	11.3%	—	0	1
Visit History					
Number of prior primary care visits	Number of primary care visits in the year prior to cohort entry	5.2	4.8	0	38
Receipt of hepatology care	History of hepatology care in the year prior to cohort entry, coded as 1 if the patient received the care, 0 otherwise.	25.7%	—	0	1

(continued)

Table 2. (continued)

Variable	Definition	Mean	SD	Min	Max
Health System Accessibility					
Insurance coverage					
Commercial	Yes = 1, otherwise = 0	3.0%	—	0	1
Medicaid	Yes = 1, otherwise = 0	20.6%	—	0	1
Medical assistance/ charity	Yes = 1, otherwise = 0	41.0%	—	0	1
Medicare	Yes = 1, otherwise = 0	24.7%	—	0	1
Self-pay	Yes = 1, otherwise = 0	2.0%	—	0	1
Unknown	Yes = 1, otherwise = 0	8.7%	—	0	1
Proximity to clinics	Whether a patient has a close geographical proximity to clinics. Coded as 1 if there are more than three clinics in the zip code that match the first three digits of the zip code where the patient resides, 0 otherwise.	66.7%	—	0	1
Neighborhood Socioeconomic Status[a]					
Educational attainment (%)	Percentage of people who are 18 years and over and received a bachelor's, master's, professional, or doctorate degree	33.6	6.6	13.8	52.9
Income ($)	Per capita income: mean income computed for every man, woman, and child in the same zip code	35,223.8	4,117.7	15,839.6	42,925.3
Average commute time (minutes)	Mean travel time to work from home during the reference week	27.1	2.1	19.1	33.6
Private health insurance coverage (%)	Percentage of civilian noninstitutionalized population with the insurance coverage provided through an employer or union, a plan purchased by an individual from a private company, or military health care.	58.6	7.2	35.3	74.8
Public health insurance coverage (%)	Percentage of civilian noninstitutionalized population with the insurance coverage provided through the federal programs Medicare, Medicaid, and Veterans Affairs Health Care, as well as the Children's Health Insurance Program and individual state health plans.	28.6	4.2	19.9	39.7
Unemployment rate (%)	Percentage of civilians 16 years old and over classified as unemployed	4.0	.4	2.5	5.7
Population	Number of people 16 years and over	1,148,050	371,086	10,824	1,791,015

[a]We used zip code to identify the neighborhoods. All zip-code-level covariates are aggregated to the level of the first three digits by calculating the sum (i.e., population) or mean (bachelor's degree or higher, mean travel time to work, and per capita income) across all five-digit zip codes that share the same first three digits.
Notes: After we exclude patients who were diagnosed with HCC or deceased, the screening completion rate is 38.5% in Period 2 and 35.4% in Period 3.

randomly assigned to one of the experimental conditions, the treatment assignment W_i is independent of the potential outcomes conditional on X_i (i.e., $\{Y_{i1}, Y_{i0}\} \perp W_i \mid X_i$). This assumption implies that the treatment is as good as random within each subpopulation indexed by $X_i = x$. Thus, given the data (X_i, Y_i, W_i), we can revise Equation 1 to the following:

$$\tau(x) = E[Y_{i1} - Y_{i0} | X_i = x]$$
$$= E[Y_i | W_i = 1, X_i = x] - E[Y_i | W_i = 0, X_i = x].$$
(2)

Common approaches to estimate the function $\tau(x)$ include nearest neighbor matching and kernel methods, but these methods do not perform well in the presence of many covariates or complex interactions among covariates (Wager and Athey 2018).

Causal forests. Causal forests combine causal inference in economics with random forests in machine learning. Random forests (Breiman 2001) deploy supervised machine learning algorithms to achieve high out-of-sample prediction accuracy with very little tuning, particularly with high dimensional data with underlying nonlinear relationships (Hastie, Tibshirani, and Friedman 2009).

Random forests (1) build a large collection of individual decision trees such that each tree predicts the outcome variable given the vector of covariates and (2) average the predictions from those trees. First, each tree is trained on a bootstrap training sample (not on the original sample) with a randomly chosen subset of covariates (not with all the covariates), and it is built by recursively partitioning the chosen covariate space into splits, determining each split by minimizing the mean squared error of the prediction of outcomes in the case of regression trees. Given the tree split, each tree clusters the most similar observations into a terminal node known as a leaf. To predict the outcome of an observation outside of the estimation sample, each tree makes a prediction using the mean of outcomes in the leaf where this new observation belongs. Finally, a random forest averages the prediction from those trees.

Researchers have recently adapted random forests to draw inferences. The technique known as causal forests utilizes an algorithm for flexible modeling of interactions in high dimensions by building many causal trees and averaging their predictions to estimate the treatment effect function $\tau(x)$. Causal forests provide valid asymptotic confidence intervals for the treatment effects (Wager and Athey 2018).
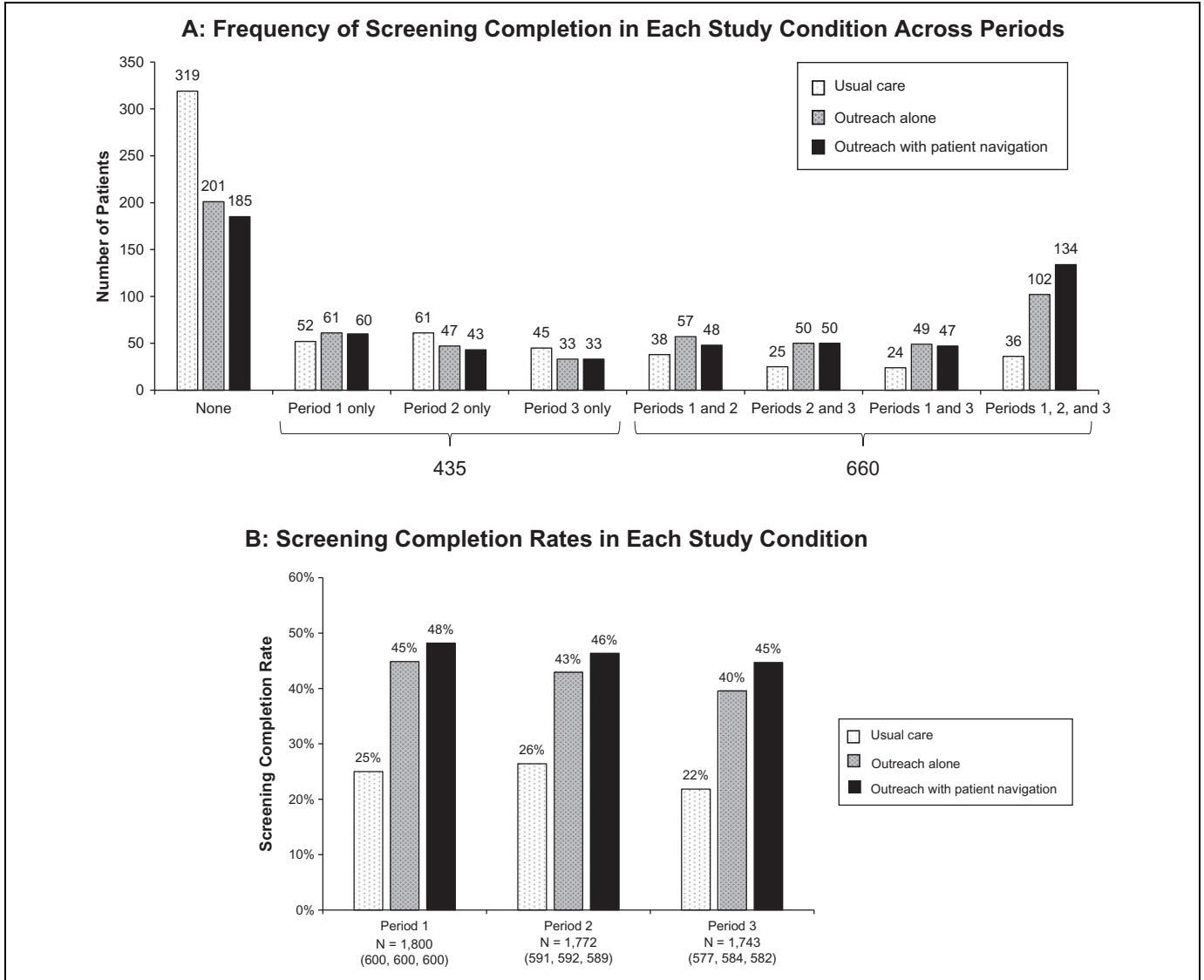
A: Frequency of Screening Completion in Each Study Condition Across Periods

Figure 2. Model-free evidence.

Given a profile of patient characteristics x, tree-based models help identify the most similar patients locally in the patient characteristics space with an adaptive neighborhood metric (i.e., similar patients are in the same leaf). Wager and Athey (2018) adapt the regression tree to estimate the within-leaf treatment effects by taking the difference between the mean outcomes of treated and control units in the same leaf:

$$\hat{\tau}(x) = \frac{1}{|\{i : W_i = 1, X_i \in L\}|} \sum_{\{i:W_i=1, X_i \in L\}}$$

$$Y_i - \frac{1}{|\{i : W_i = 0, X_i \in L\}|} \sum_{\{i:W_i=0, X_i \in L\}} Y_i. \quad (3)$$

To ensure consistency and asymptotic normality, Wager and Athey (2018) prove a bias-reducing condition called honesty: a

tree achieves honesty if each bootstrap training sample only uses the outcome of interest Y_i to estimate the within-leaf treatment effect based on Equation 3 or to determine where to split the covariate space, but not both. In other words, the bootstrap training sample is further split into two subsamples: one used to build the tree (i.e., understand where the treatment heterogeneity is given the vector of covariates),[10] and the other used to estimate the treatment effects given the tree structure.

---

[10] Unlike regression trees, where the splits are determined by minimizing mean squared error of the prediction of outcomes, causal trees are built by minimizing the expected mean squared error of predicted treatment effects, which is equivalent to maximizing the variance of treatment effects across leaves minus a penalty for within-leaf variance (Athey and Imbens 2016).

Using this process, causal forests produce an ensemble of B such trees (Breiman 2001; Wager and Athey 2018), each of which outputs an estimate $\hat{\tau}_b(x)$ and averages the predictions from those trees to compute an estimated CATE: $\hat{\tau}(x) = B^{-1}\sum_{b=1}^{B}\hat{\tau}_b(x)$.

This aggregation scheme also helps reduce variance and smooths sharp decision boundaries (Bühlmann and Yu 2002). The variance estimate of causal forests is defined as follows (Efron 2014; Wager and Athey 2018; Wager, Hastie, and Efron 2014):

$$\hat{V}(x) = \frac{n-1}{n}\left(\frac{n}{n-s}\right)^2 B^{-1}\sum_{i=1}^{n} COV[\hat{\tau}_b(x), N_{ib}]^2, \quad (4)$$

where $\hat{\tau}_b(x)$ is the treatment effect estimate from the bth tree. $N_{ib} \in \{0, 1\}$ indicates whether the bootstrap training sample i is used for the tree b, $n(n-1)/(n-s)^2$ is a finite-sample correction for forests grown by subsampling without replacement, and the covariance is taken with respect to all B trees in the forest. Equations 3 and 4 produce a treatment effect estimate and a confidence interval for each patient.

In marketing, causal forests have been applied in the context of customer retention (Ascarza 2018), information disclosure and physician payments (Guo, Sriram, and Manchanda 2019), and adoption of voice-activated shopping devices and consumers' purchase quantity, spending amount, and search activities (Sun et al. 2019). To our knowledge, this is the first study to use causal forests in the context of randomized health care field experiments.

Application to our context. Following Singal et al. (2019), we have two different treatment conditions (outreach alone and outreach with patient navigation) and three different periods. We use the following procedure to perform six causal forest estimations. Additional aspects of the estimation are summarized in Web Appendix D3.

- *Step 1.* Using patient characteristics as covariates,[11] we applied causal forests to obtain each patient's treatment effect estimate in the sample that includes patients in the baseline (condition 1, sample size = 600) and those in the outreach-alone condition (condition 2, sample size = 600) in Period 1. For each patient i in condition 1 in Period 1, the patient-level treatment effect estimate is $\hat{\tau}_{i,P1,1\to2}^{1}$ (i.e., the difference between the outcome we observe for the patient i in condition 1 and the outcome that would be realized if this patient were in condition 2); for each patient i in condition 2 in Period 1, the patient-level treatment effect estimate is $\hat{\tau}_{i,P1,2\to1}^{2}$ (i.e., the difference between the outcome we observe for the patient i in condition 2 and the outcome that would be realized if this patient were in condition 1). We term this first causal forest estimation Forest$_{P1}^{12}$, where P1 refers

to Period 1, and the superscript 12 refers to the comparison of the baseline condition (1) and the outreach-alone condition (2).

- *Step 2.* After excluding the patients who were deceased or diagnosed with HCC in the previous period(s), we repeated Step 1 to obtain $\hat{\tau}_{i,P2,1\to2}^{1}$ and $\hat{\tau}_{i,P2,2\to1}^{2}$ (condition 1, sample size = 591; condition 2, sample size = 592) in Period 2 and $\hat{\tau}_{i,P3,1\to2}^{1}$ and $\hat{\tau}_{i,P3,2\to1}^{2}$ (condition 1, sample size = 577; condition 2, sample size = 584) in Period 3. As discussed, we included one (two) additional covariate(s) indicating whether a patient has completed the screening test in the prior period(s) in the causal forest estimation of Period 2 (3). We term these second and third causal forests Forest$_{P2}^{12}$ and Forest$_{P3}^{12}$, where P2 and P3 refer to Period 2 and Period 3, respectively, and the superscript 12 refers to the comparison of the baseline condition (1) and the outreach-alone condition (2).

- *Step 3.* We repeated Step 1 to obtain each patient's treatment effect estimate in the sample that includes patients in the baseline (condition 1, sample size = 600) and those in the outreach-with-patient-navigation condition (condition 3, sample size = 600) in Period 1. For each patient in condition 1, the patient-level treatment effect estimate is $\hat{\tau}_{i,P1,1\to3}^{1}$ (i.e., the difference between the outcome we observe for the patient i in condition 1 and the outcome that would be realized if this patient were in condition 3); for each patient in condition 3, the patient-level treatment effect estimate is $\hat{\tau}_{i,P1,3\to1}^{3}$ (i.e., the difference between the outcome we observe for the patient i in condition 3 and the outcome that would be realized if this patient were in condition 1). We term this fourth causal forest Forest$_{P1}^{13}$, where P1 refers to Period 1, and the superscript 13 refers to the comparison of the baseline condition (1) and outreach-with-patient-navigation condition (3).

- *Step 4.* We repeated Step 2 to obtain $\hat{\tau}_{i,P2,1\to3}^{1}$ and $\hat{\tau}_{i,P2,3\to1}^{3}$ in Period 2 (condition 1, sample size = 591; condition 3, sample size = 589) and $\hat{\tau}_{i,P3,1\to3}^{1}$ and $\hat{\tau}_{i,P3,3\to1}^{3}$ (condition 1, sample size = 577; condition 3, sample size = 582) in Period 3. We term these fifth and sixth causal forests Forest$_{P2}^{13}$ and Forest$_{P3}^{13}$, where P2 and P3 refer to Period 2 and Period 3, respectively, and the superscript 13 refers to the comparison of the baseline condition (1) and the outreach-with-patient-navigation condition (3).

Table 3 and Figure 3 show the distribution of the patient-level treatment effect estimates based on the causal forest estimation. Relative to the baseline condition, outreach alone (outreach with patient navigation) increases screening completion rate by between 10 and 20 (13 and 24) percentage points (Table 3, Panel A). Causal forests enable us to construct confidence intervals for patient-level treatment effect estimates. As we report in Table 3, Panel B, outreach-alone intervention induces positive and statistically significant treatment effects among 100%, 74%, and 66% of the patients in Periods 1, 2, and 3,

---

[11] We scale continuous variables to zero mean and unit variance and expand categorical variables via one-hot encoding.

Table 3. Summary of Average Treatment Effects (ATEs) and Patient-Level Conditional Average Treatment Effects (CATEs) by Outreach Type.

A: ATEs by Outreach Type

	Outreach Alone				Outreach with Patient Navigation				
	N	ATE	SE	95% CI		N	ATE	SE	95% CI	
Period 1 ATE	1,200	.200	.026	.148	.251	1,200	.235	.026	.184	.287
Period 2 ATE	1,183	.109	.027	.057	.162	1,180	.128	.028	.074	.182
Period 3 ATE	1,161	.101	.025	.052	.150	1,159	.128	.026	.077	.178

B: Patient-Level CATEs by Outreach Type

	Outreach Alone					Outreach with Patient Navigation				
	N	Mean	SD	Min	Max	N	Mean	SD	Min	Max
Period 1										
Patient-level CATEs	1,200	.199	.036	.112	.311	1,200	.236	.046	.119	.366
Significant patient-level CATEs[a]	1,200	.199	.036	.112	.311	1,200	.236	.046	.119	.366
Proportion of significant patient-level CATEs			100%					100%		
Period 2										
Patient-level CATEs	1,183	.108	.038	−.006	.192	1,180	.127	.042	.031	.230
Significant patient-level CATEs[a]	875	.125	.024	.057	.192	975	.139	.036	.052	.230
Proportion of significant patient-level CATEs			74%					83%		
Period 3										
Patient-level CATEs	1,161	.099	.036	.014	.192	1,159	.126	.030	.047	.207
Significant patient-level CATEs[a]	767	.118	.028	.047	.192	1,030	.131	.027	.062	.207
Proportion of significant patient-level CATEs			66%					89%		

[a]Statistical significance is at the 95% level.
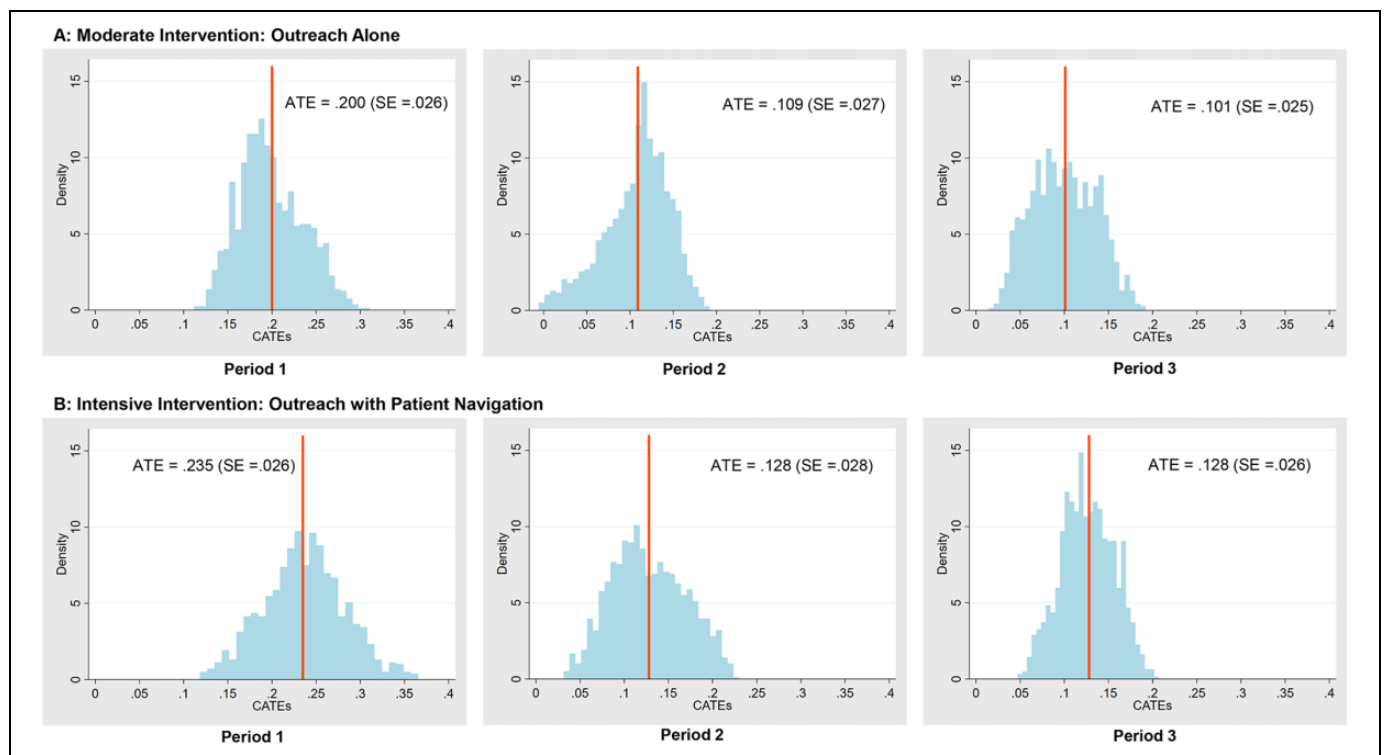*Notes:* CI = confidence interval.

Figure 3. Distribution of patient-level CATEs.
Notes: ATE refers to the average treatment effect; SE refers to the standard error; CATEs refer to conditional average treatment effects.

respectively ($p < .05$), while the outreach-with-patient-navigation intervention does so among 100%, 83%, and 89% of the patients in Periods 1, 2, and 3, respectively ($p < .05$).

Germane to the focus of this article, there is substantial heterogeneity in those significant patient-level treatment effects: (1) compared with outreach alone, outreach-with-patient-navigation intervention induces a higher proportion of patients with significant positive treatment effect estimates in Periods 2 ($83\% - 74\% = 9\%$) and 3 ($89\% - 66\% = 23\%$), and (2) patient-level treatment effect estimates of outreach-alone (outreach-with-patient-navigation) intervention range from 5–31 (5–37) percentage points. Next, we investigate the sources of heterogeneity.

Incorporating Heterogeneity in Patient-Level Treatment Effects

We examine the treatment effect heterogeneity by correlating treatment effect estimates with patient characteristics. Accordingly, we estimated the following equations:

$$
\begin{aligned}
\hat{\tau}^2_{ijt=\{P1,P2,P3\}} = \ & \alpha_{20} + \alpha_{21}\,\text{Age}_{ij} + \alpha_{22}\,\text{Gender}_{ij} \\
& + \alpha_{23}\,\text{Ethnicity}_{ij} + \alpha_{24}\,\text{Language}_{ij} \\
& + \alpha_{25}\,\text{Child-PughB}_{ij} + \alpha_{26}\,\text{Charlson}_{ij} \\
& + \alpha_{27}\,\text{Cirrhosis}_{ij} + \alpha_{28}\,\text{Etiology}_{ij} \\
& + \alpha_{29}\,\text{Prior\_visit}_{ij} + \alpha_{210}\,\text{Hepatology\_care}_{ij} \\
& + \alpha_{211}\,\text{Insurance Coverage}_{ij} + \alpha_{212}\,\text{Proximity}_{ij} \\
& + \alpha_{213}\,\text{Education}_j + \alpha_{214}\,\text{Income}_j \\
& + \alpha_{215}\,\text{Commute}_j + \alpha_{216}\,\text{Private}_j \\
& + \alpha_{217}\,\text{Public}_j + \alpha_{218}\,\text{Unemploy}_j \\
& + \alpha_{219}\,\text{Population}_j + \eta_t + \epsilon_{2ijt},
\end{aligned}
$$
(5a)

$$
\begin{aligned}
\hat{\tau}^3_{ijt=\{P1,P2,P3\}} = \ & \alpha_{30} + \alpha_{31}\,\text{Age}_{ij} + \alpha_{32}\,\text{Gender}_{ij} \\
& + \alpha_{33}\,\text{Ethnicity}_{ij} + \alpha_{34}\,\text{Language}_{ij} \\
& + \alpha_{35}\,\text{Child-PughB}_{ij} + \alpha_{36}\,\text{Charlson}_{ij} \\
& + \alpha_{37}\,\text{Cirrhosis}_{ij} + \alpha_{38}\,\text{Etiology}_{ij} \\
& + \alpha_{39}\,\text{Prior\_visit}_{ij} + \alpha_{310}\,\text{Hepatology\_care}_{ij} \\
& + \alpha_{311}\,\text{Insurance Coverage}_{ij} + \alpha_{312}\,\text{Proximity}_{ij} \\
& + \alpha_{313}\,\text{Education}_j + \alpha_{314}\,\text{Income}_j \\
& + \alpha_{315}\,\text{Commute}_j + \alpha_{316}\,\text{Private}_j \\
& + \alpha_{317}\,\text{Public}_j + \alpha_{318}\,\text{Unemploy}_j \\
& + \alpha_{319}\,\text{Population}_j + \eta_t + \epsilon_{3ijt},
\end{aligned}
$$
(5b)

where $\hat{\tau}^2_{ijt=\{P1,P2,P3\}}$ and $\hat{\tau}^3_{ijt=\{P1,P2,P3\}}$ refer to patient-level treatment effect estimates of outreach alone and those of outreach with patient navigation, j denotes the three-digit zip code, and t denotes the period. We pooled the estimates across periods and included period-fixed effects (η_t) to capture common time-varying observables that may affect them and clustered standard errors at the patient level to allow for heteroskedasticity and correlated errors within patients over time.

Columns 1 and 2 of Table 4 report the results showing the sources of heterogeneity in patient-level treatment effects. Next, we discuss the patient characteristics associated with the treatment effect heterogeneity.

Age. Older patients are less responsive to the outreach-alone intervention than younger patients ($\hat{\alpha}_{21} = -.003, p < .001$) but they are more responsive to outreach-with-patient-navigation intervention than younger patients ($\hat{\alpha}_{31} = .006, p < .001$). A possible explanation is that older adults prefer to use information that is customized to their needs rather than generic information that can be overwhelming (Cole et al. 2008), which makes them less responsive to direct mails than younger adults (Kaldenberg, Koenig, and Becker 1994). The interactive and personalized nature of the navigation over the telephone provides targeted and useful information to older patients, making it more effective (King, Rejeski, and Buchner 1998).

Gender. Female patients are more responsive to both outreach interventions than male patients (outreach alone: $\hat{\alpha}_{22} = .014, p < .001$; outreach with patient navigation: $\hat{\alpha}_{32} = .007, p < .001$). This is likely due to the higher prevention and loss-minimization focus among women (Trudeau et al. 2003). According to agency-communion theory (Carlson 1971), men focus on maximizing gains while women focus on minimizing the downside potential of their decision. Outreach messages for cancer screening, by design, approach health care from a prevention and loss-minimization focus.

Ethnicity and language. Hispanic patients are more responsive to both outreach interventions than Caucasian patients (outreach alone: $\hat{\alpha}_{23}^H = .011, p < .001$; outreach with patient navigation: $\hat{\alpha}_{33}^H = .004, p < .001$). Likewise, non-Hispanic African American patients are more responsive to both outreach interventions than Caucasian patients (outreach alone: $\hat{\alpha}_{23}^{AA} = .010, p < .001$; outreach with patient navigation: $\hat{\alpha}_{33}^{AA} = .011, p < .001$). Similarly, patients whose primary language is Spanish are more responsive to both outreach interventions than those whose primary language is English (outreach alone: $\hat{\alpha}_{24} = .014, p < .001$; outreach with patient navigation: $\hat{\alpha}_{34} = .015, p < .001$). Due to language and access barriers, such patients may have relatively fewer opportunities to learn about the health screening information than ethnic majority groups (Caplan, Wells, and Haynes 1992; Szczepura 2005). Given the lower baseline access, outreach interventions that provide information on screening opportunities should be more effective among such groups than their counterparts (Lasser et al. 2011).

Health status. Patients in a poorer health status (those with Child-Pugh B) are less responsive to both outreach interventions than patients with a better health status (outreach alone: $\hat{\alpha}_{25} = -.012, p < .001$; outreach with patient navigation: $\hat{\alpha}_{35} = -.003, p < .001$). The pattern is consistent when Charlson Comorbidity Index and the presence of documented

Table 4. Sources of Heterogeneity in Patient-Level CATEs.

	(1) Outreach Alone		(2) Outreach with Patient Navigation		(3) Outreach Alone		(4) Outreach with Patient Navigation	
	Est	SE	Est	SE	Est	SE	Est	SE
Demographics								
Age	−.003***	(.000)	.006***	(.000)	−.003***	(.000)	.005***	(.000)
Gender (female = 1)	.014***	(.001)	.007***	(.001)	.013***	(.001)	.007***	(.001)
Hispanic	.011***	(.001)	.004***	(.001)	.012***	(.001)	.005***	(.001)
Non-Hispanic African American	.010***	(.001)	.011***	(.001)	.010***	(.001)	.011***	(.001)
Other/unknown	.009***	(.002)	.007*	(.003)	.008***	(.002)	.007*	(.003)
Spanish	.014***	(.001)	.015***	(.001)	.013***	(.001)	.014***	(.001)
Other	−.002	(.002)	.003	(.005)	−.001	(.004)	.004	(.005)
Health Status								
Child-Pugh B	−.012***	(.001)	−.003***	(.001)	−.011***	(.001)	−.003***	(.001)
Charlson Comorbidity Index	−.003***	(.000)	−.002***	(.000)	−.004***	(.000)	−.003***	(.000)
Documented cirrhosis	−.002*	(.001)	−.002*	(.001)	−.002**	(.001)	−.003**	(.001)
Hepatitis B	−.002*	(.001)	.001	(.001)	−.001	(.001)	.002	(.001)
Alcohol-induced	−.004*	(.002)	−.001	(.003)	−.005**	(.001)	.000	(.002)
Nonalcoholic steatohepatitis	−.003***	(.001)	−.000	(.001)	−.003***	(.001)	.000	(.001)
Other	−.000	(.001)	.005***	(.001)	−.001	(.001)	.004***	(.001)
Visit History								
Number of prior primary care visits	.010***	(.001)	.011***	(.001)	.020***	(.001)	.022***	(.002)
Receipt of hepatology care	.003***	(.001)	.003*	(.001)	.002*	(.001)	.001	(.001)
Health System Accessibility								
Commercial	−.009***	(.002)	−.005*	(.002)	−.009***	(.001)	−.002	(.002)
Medicaid	−.010***	(.001)	−.001	(.001)	−.009***	(.001)	.001	(.001)
Medicare	−.016***	(.001)	−.002*	(.001)	−.015***	(.001)	−.002*	(.001)
Self-pay	−.012***	(.002)	−.003	(.003)	−.012***	(.002)	−.002	(.002)
Unknown	−.013***	(.001)	−.008***	(.001)	−.011***	(.001)	−.005**	(.001)
Proximity to clinics	.008***	(.001)	.005**	(.002)	.008***	(.001)	.007***	(.001)
Neighborhood Socioeconomic Status								
Educational attainment (%)	.002	(.004)	−.011	(.006)	.003	(.004)	−.011	(.006)
Income ($)	−.000	(.003)	.012**	(.004)	.000	(.003)	.011**	(.004)
Average commute time (minutes)	−.004***	(.001)	−.000	(.002)	−.003**	(.001)	−.002	(.001)
Private health insurance (%)	−.000	(.004)	.002	(.005)	−.002	(.003)	.008	(.005)
Public coverage (%)	.011**	(.004)	.006	(.005)	.009**	(.003)	.009*	(.004)
Unemployment rate (%)	.002	(.001)	.003	(.002)	.002	(.001)	.005	(.002)
Population	.005*	(.002)	.005*	(.002)	.003	(.002)	.004*	(.002)
Period Fixed Effects								
Period 2 dummy	−.091***	(.001)	−.109***	(.001)	−.091***	(.001)	−.109***	(.001)
Period 3 dummy	−.100***	(.001)	−.110***	(.002)	−.100***	(.001)	−.110***	(.002)
Exploratory Interactions								
Primary care visit^2					−.003***	(.000)	−.004***	(.000)
Primary care visit × Age					.000	(.000)	.000	(.001)
Primary care visit × Gender					.000	(.001)	−.003**	(.001)
Primary care visit × Hispanic					−.003**	(.001)	−.002	(.002)
Primary care visit × African American					.001	(.001)	−.001	(.001)
Primary care visit × Spanish					−.005***	(.001)	−.005***	(.002)
Primary care visit × Child-Pugh B					−.003***	(.001)	−.003**	(.001)
Primary care visit × Charlson Index					−.000	(.000)	−.001	(.000)
Primary care visit × Medicaid					−.000	(.001)	.003	(.002)
Primary care visit × Medical assistance					−.002*	(.001)	−.001	(.002)
Primary care visit × Medicare					−.001	(.001)	−.001	(.002)
Primary care visit × Proximity to clinics					−.000	(.001)	.002	(.001)
Intercept	.190***	(.001)	.225***	(.002)	.193***	(.001)	.227***	(.002)
Clustered standard error	Yes		Yes		Yes		Yes	
R^2	.770		.727		.783		.741	
N	3,544		3,539		3,544		3,539	

*$p < .05$.
**$p < .01$.
***$p < .001$.
Notes: The baseline categories of main effects are male, non-Hispanic Caucasian, Hepatitis C, English, medical assistance/charity, and Period 1. We scaled continuous variables to zero mean and unit variance. As we pooled the estimates of three periods, sample sizes are 3,544 (1,200 + 1,183 + 1,161) and 3,539 (1,200 + 1,180 + 1,159) (see sample size columns of Table 3, Panel B).

cirrhosis are used as indicators of health status. A possible explanation is that outreach interventions might make patients fearful of finding out they have cancer (Aro et al. 2001) and experience death anxiety (Grossman et al. 2018). Those with poor health will experience higher death anxiety due to lower optimism about their health (Arndt, Routledge, and Goldenberg 2006), which reduces adaptive coping and thus decreases the utilization of health care services (Moorman and Matulich 1993). We also find that compared with patients with Hepatitis C, those with Hepatitis B are less responsive to outreach alone (coefficients ranging from −.004 to −.002), but this is not the case for outreach with patient navigation.

Visit history. Patients with a higher number of prior primary care visits are more responsive to both outreach interventions than those with fewer prior primary care visits (outreach alone: $\hat{\alpha}_{29}$ = .010, $p < .001$; outreach with patient navigation: $\hat{\alpha}_{39}$ = .011 $p < .001$). Similarly, patients who previously received hepatology care are more responsive to both outreach interventions than patients with no prior hepatology care (outreach alone: $\hat{\alpha}_{210}$ = .003, $p < .001$; outreach with patient navigation: $\hat{\alpha}_{310}$ = .003, $p < .05$). At its core, a patient's prior visit history signifies the extent to which a patient has a favorable attitude toward utilizing health care services to pursue their health goals (Klein and Cerully 2007) and has familiarity with the utilization process (Goldman et al. 2015). This should motivate patients to get screened.

Health system accessibility. Patients with insurance coverage through medical assistance/charity are generally more responsive to both outreach interventions than patients with other types of insurance (outreach alone: $\hat{\alpha}_{211}$ = ranging from −.016 to −.009, $p < .001$; outreach with patient navigation: $\hat{\alpha}_{211}$ = ranging from −.008 to −.001, $p < .001$ through n.s.). Patients who receive health care at a low cost due to medical assistance/charity, with access to the corresponding insurance, are more likely to respond to outreach interventions because of their ability to overcome financial hardships to utilize screening services. A patient's ease of accessing health care services is based on not only their ability to pay for the service but also their proximity to health care providers. We find that patients with closer proximity to care are more responsive to both outreach interventions than patients with further proximity to care (outreach alone: $\hat{\alpha}_{212}$ = .008, $p < .001$; outreach with patient navigation: $\hat{\alpha}_{312}$ = .005, $p < .01$).

Neighborhood socioeconomic status. Patients who live in more educated neighborhoods are not necessarily more or less responsive to interventions (outreach alone: $\hat{\alpha}_{213}$ = .002, n.s.; outreach with patient navigation: $\hat{\alpha}_{313}$ = −.011, n.s.). Yet patients who reside in a higher-income neighborhoods are more responsive to outreach with patient navigation (outreach alone: $\hat{\alpha}_{214}$ = −.000, n.s.; outreach with patient navigation: $\hat{\alpha}_{314}$ = .012, $p < .01$), which implies that those in low-income neighborhoods are less responsive to this intervention. Patients in low-income neighborhoods face unique challenges such as

higher rates of obesity, chronic disease, environmental pollutants, and incarceration (Khullar and Chokshi 2018). The prevalent health and environmental challenges in these communities might cause anxiety among community members and lead them to be pessimistic about their health (Conger et al. 1992), thus making them less responsive to outreach intervention. Patients in neighborhoods with longer average commute times are less responsive to outreach alone (outreach alone: $\hat{\alpha}_{215}$ = −.004, $p < .001$), but this is no longer the case for outreach with patient navigation (outreach with patient navigation: $\hat{\alpha}_{315}$ = −.000, n.s.). Patient navigation alleviates perceived costs associated with a screening by providing the information on the estimated duration for the appointment, so patients who live in a highly trafficked community will no longer show resistance to a screening.

While patient-level insurance coverage should capture the impact of health system accessibility, the neighborhood-level health insurance coverage can also offer additional insights. Patients' responsiveness to the outreach interventions does not vary by the degree of private health insurance coverage in their neighborhood (outreach alone: $\hat{\alpha}_{216}$ = −.000, n.s.; outreach with patient navigation: $\hat{\alpha}_{316}$ = .002, n.s.). However, patients in a neighborhood with a greater public health insurance coverage are more responsive to the outreach-alone intervention but not to the outreach-with-patient-navigation intervention (outreach alone: $\hat{\alpha}_{217}$ = .011, $p < .01$; outreach with patient navigation: $\hat{\alpha}_{317}$ = .006, n.s.).

Neighborhood unemployment rate does not significantly affect patients' responsiveness to the interventions (outreach alone: $\hat{\alpha}_{218}$ = .002, n.s.; outreach with patient navigation: $\hat{\alpha}_{318}$ = .003, n.s.). Yet patients from neighborhoods with more dense populations are more responsive to both interventions (outreach alone: $\hat{\alpha}_{319}$ = .005, $p < .05$; outreach with patient navigation: $\hat{\alpha}_{219}$ = .005, $p < .05$), implying that patients in rural areas are less responsive to interventions. These results bear a notable caveat: a potential aggregation bias due to the measurement of neighborhood variables at the three-digit-zip-code level may have led to the null effect.

Additional post hoc analysis combining patient characteristics. As described in Columns 3 and 4 of Table 4, we explore possible combinations of patient characteristics with the interactions between prior primary care visits and other patient characteristics. This is akin to examining higher-order interactions in an analysis of variance. This analysis offers several insights.

The marginal benefit of additional primary care visits diminishes such that a patient's primary care visit has a nonlinear effect on outreach intervention effectiveness. Referring to Columns 3 and 4 of Table 4, there is a positive linear coefficient (outreach alone: b = .020, $p < .001$; outreach with patient navigation: b = .022, $p < .001$) and a negative quadratic coefficient (outreach alone: b^2 = −.003, $p < .001$; outreach with patient navigation: b^2 = −.004, $p < .001$) for the effect. Jointly, the coefficients capture diminishing returns such that a patient's first few primary care visits yield large marginal returns. Given the importance of the initial visits, health care

professionals can enhance outreach effectiveness by targeting patients who have made fewer than numerous visits in the past.

The interactions between prior primary care visits and patient characteristics can help practitioners further identify the responsive subgroups. For example, Spanish-speaking patients' responsiveness to outreach interventions is attenuated as primary care visits increase (coefficient $= -.005\ p < .001$ for both outreach alone and outreach with patient navigation). It could be that Spanish-speaking patients perceive that outreach interventions are less informative than primary care visits. Practitioners may target Spanish-speaking patients who have no prior primary care visits in the past. Patients with Child-Pugh B are even less responsive to outreach interventions as primary care visits increase (outreach alone: coefficient $= -.003\ p < .001$; outreach with patient navigation: coefficient $= -.003, p < .01$), suggesting that increased primary care visits compound the perception of a fear and death anxiety of having cancer triggered by outreach interventions and thus decrease the use of screening. Overall, our post hoc analysis highlights the need to understand how outreach effectiveness varies by the combination of patient characteristics.

Dynamics in the Proportion of Treatment Effects that Are Statistically Significant

The dynamics in the proportion of treatment effects that are statistically significant across each condition display an interesting pattern. As Figure E1 in Web Appendix E shows, there is very little heterogeneity in Period 1 in terms of the proportion of treatment effects that are statistically significant. Specifically, 100% (100%) of the treatment effects due to outreach alone (outreach with patient navigation) are statistically significant in Period 1. However, compared with outreach alone, outreach with patient navigation induces a higher proportion of patients with significant positive treatment effect estimates in Period 2 (83% vs. 74%) and Period 3 (89% vs. 66%). The effectiveness of the outreach with patient navigation relative to outreach alone improves over time. Medical and health care professionals believe that exposure to outreach nurtures the motivation for screening compliance, which is the crucial first step. Once patients are compliant to screening, potential barriers to getting screening may be addressed in subsequent periods. Thus, it is possible that the heterogeneity in the proportion of significant treatment effects could be related to the repetition of the treatment over different periods.

As a thought experiment, we conducted an analysis investigating the extent of heterogeneity in the proportion of treatment effects that are statistically significant when we "turn off" the repeated nature of the treatment. In this analysis, we define the dependent variable as whether a patient undergoes a screening at least once in the short (0–6 months), medium (0–12 months), and long (0–18 months) runs. The goal of this approach is to study whether interventions can bring at-risk patients in for a screening at least once during the three periods investigated. We also wanted to understand if heterogeneity, in terms of proportion of treatment effects that are statistically significant, manifests when we "turn off" the repeated nature of the treatment. Web Appendix Figure E2 reports the screening completion rates and plots the patient-level treatment effects, and Table E2 reports the summary of the treatment effects across each condition. The main takeaways are as follows:

- *There is little heterogeneity in the proportion of treatment effects that are statistically significant.* Web Appendix Table E2 reports that outreach-alone intervention induces positive and statistically significant treatment effects among 100%, 100%, and 99.9% of the patients in Periods 1, 2, and 3, respectively ($p < .05$), while the outreach-with-patient-navigation intervention does so among 100%, 99.7%, and 98.5% of the patients in Periods 1, 2, and 3, respectively ($p < .05$).
- *There remains substantial heterogeneity in the magnitude of treatment effects.* Web Appendix Table E2 reports that patient-level treatment effect estimates of outreach-alone (outreach-with-patient-navigation) intervention range from 10–32 (11–37) percentage points.

Web Appendix Figures E1, E2, and Table E2 jointly show that the heterogeneity in the proportion of treatment effects that are statistically significant is related to the dynamics induced by the repeated nature of the treatment. Future research should investigate the sources of these dynamics.

Return on Cancer Outreach Interventions

We evaluate the return on outreach interventions among patients and across three periods:

$$
\begin{aligned}
\text{Return}_k = \sum_{t=1}^{3} \sum_{i=1}^{N_{t=\{1,2,3\}}} & \{\Pr(\text{Screening}_{ikt} = 1) \\
& \times [\text{Benefit}_{ikt} - \text{Screening Cost}_{ikt} \\
& - \Pr(\text{Early Tumor}_{ikt} = 1 | \text{Screening}_{ikt} = 1) \\
& \times \text{Treatment Cost}_{ikt}] - \Pr(\text{Screening}_{ikt} = 0) \\
& \times \text{Opportunity Cost}_{ikt} - \text{Outreach Cost}_{ikt} \},
\end{aligned}
$$
(6)

where $\Pr(\text{Screening}_{ikt} = 1)$ refers to the probability that patient i assigned to outreach type k completes the screening in period t.

- If a patient completes the screening test:
  - The health care institution generates Benefit$_{ikt}$ for patient i receiving intervention type k in period t, captured by the quality-adjusted life years of a patient attributable to the screening (typically expressed in the financial value in the medical literature).
  - The health care institution incurs Screening Cost$_{ikt}$ for patient i receiving intervention type k in period t, which includes the costs of an ultrasound/MRI/CT test or a combination of these tests (i.e., each patient can complete multiple tests).

○ Conditional on being detected with an early tumor, the health care provider incurs Treatment Cost$_{ikt}$ for patient i intervention type k in period t, which includes the costs of tumor resection, liver transplantation, and local ablative therapies.

● If a patient does not complete the screening test:
  ○ The health care institution incurs Opportunity Cost$_{ikt}$ if patient i receiving intervention type k in period t develops advanced HCC, which creates costs.

● Irrespective of whether the patient completes the screening test:
  ○ Outreach Cost$_{ikt}$ is incurred if the health care institution employs an outreach program. The outreach costs are higher for the outreach-with-patient-navigation than the outreach-alone condition and are zero for the baseline condition.

Research has documented that HCC screening completion with biannual ultrasound extends patients' quality-adjusted life expectancy by 1.3 months, and HCC screening utilization with MRI does so by 2 months (Goossens et al. 2017). Patients may complete an ultrasound, an MRI, a CT scan, or a mix of these tests. We assume the average quality-adjusted life expectancy to be 1.65 months for each patient who completes the screening. The medical literature posits that the financial value per quality-adjusted life year is $50,000 (Andersson et al. 2008; Goossens et al. 2017). Thus, total benefits can be obtained by multiplying the number of quality-adjusted life years by the financial value per quality-adjusted life year ($50,000) (i.e., multiply total number of patients who complete the screening by average quality-adjusted life expectancy).

Table 5 presents the results of the benefit–cost calculation using Equation 6 for each condition in each period. We use the observed values in the data (e.g., actual number of patients who visit) in conjunction with parameters (e.g., early detection rate) from the medical literature to calculate the return in each condition in each period. Web Appendix F documents details on parameters from the medical literature.

No-outreach condition. We observe that 150/600, 156/591, and 126/577 patients in no-outreach condition completed the screening in Periods 1, 2, and 3, respectively. The total benefits of the no-outreach condition across the three periods are estimated to be $2,970,000.

The screening costs are the total costs of ultrasound, CT, and MRI tests completed. The number of ultrasound (CT and MRI) tests completed is 127 (69), 149 (68), and 113 (59) in Periods 1, 2, and 3. The cost per ultrasound is $143, while the average cost of CT and MRI is $1,020. Thus, the total screening costs are estimated to be $255,547. Focusing on treatment costs, 5% of the total number of screening tests typically result in early tumor detection. The average treatment cost per patient for early tumor detection is $74,397 (Goossens et al. 2017). Given

that 150/600, 156/591, and 126/577 patients in the no-outreach condition completed the screening in Periods 1, 2, and 3, 5% of them would undergo treatment costs, giving a total treatment cost of $1,606,967. Opportunity cost is incurred if patients who have not completed the screening develop advanced HCC. The annual cost of advanced HCC is $41,320, and the annual HCC probability is 2.9% (Goossens et al. 2017). Multiplying the number of patients who have not completed the screening in the no-outreach condition by the probability of HCC (2.9%) gives us a total opportunity cost of $1,600,902. Finally, outreach costs are zero in the no-outreach condition. Subtracting the total cost from the total benefit, the total return in the no-outreach condition is −$493,416, which translates to a loss of $840 per patient to the health care system.

Outreach-alone condition. Compared with the no-outreach condition, there are higher benefits in the outreach-alone case ($5,183,750 vs. $2,970,000), as there are 269/600 patients, 254/592 patients, and 231/584 patients who have completed the screening in Periods 1, 2, and 3, respectively. Using the same approach as the one used for no-outreach condition to calculate the costs, the screening costs, treatment costs, and opportunity costs, respectively, are $323,919, $2,804,752, and $1,224,642 in the outreach-alone condition. In addition, the total number of hours devoted to outreach calls is 3,477, 2,607, and 2,135. Assuming a $15 hourly wage, the total outreach cost in the outreach-alone condition is $123,285. Thus, the total return in the outreach-alone condition is $707,152, which translates to a gain of $1,192 per patient to the health care system.

Outreach-with-patient-navigation condition. Compared with the no-outreach condition, there are higher benefits in the outreach-with-patient-navigation condition ($5,651,250 vs. $2,970,000), as there are 289/600 patients, 273/589 patients, and 260/582 patients who would complete the screening. Following the same calculation approach, the screening costs, treatment costs, opportunity costs, and outreach costs are $332,871, $3,057,700, $1,137,168, and $157,548, respectively, in the outreach-with-patient-navigation condition. Thus, the total return in the outreach-with-patient-navigation condition is $965,963, which is substantially greater than that in the no-outreach condition and translates to a gain of $1,635 per patient to the health care system.

Summary of returns on nontargeted cancer outreach interventions. No outreach results in a net loss of $840 per patient to the medical hospital, whereas outreach alone (outreach with patient navigation) generates a monetary gain of $1,192 ($1,635) per patient. When extrapolated to the 3,217 patients eligible for randomization from the hospital's patient database, the cancer intervention results in a loss of $900,718 from no outreach or usual care, a gain of $1,278,402 from outreach alone, and a gain of $1,753,237 from outreach with patient navigation. In this scenario, the total gain is $2,130,921.

Table 5. Return on Nontargeted Outreach Interventions.

Cost and Benefit by Intervention Type and Across Period	Usual Care				Outreach Alone				Outreach with Patient Navigation			
	Period 1	Period 2	Period 3	Total	Period 1	Period 2	Period 3	Total	Period 1	Period 2	Period 3	Total
Sample Size	600	591	577		600	592	584		600	589	582	
Costs												
Outreach costs	$0	$0	$0	$0	$52,157	$39,099	$32,030	$123,285	$60,594	$51,338	$45,617	$157,548
Screening costs	$88,541	$90,667	$76,339	$255,547	$111,335	$118,351	$94,233	$323,919	$127,598	$105,015	$100,258	$332,871
Treatment costs	$557,975	$580,293	$468,699	$1,606,967	$1,000,634	$944,837	$859,281	$2,804,752	$1,075,031	$1,015,514	$967,156	$3,057,700
Opportunity costs	$539,226	$521,252	$540,424	$1,600,902	$396,631	$405,019	$422,993	$1,224,642	$372,665	$378,656	$385,846	$1,137,168
Total costs	$1,185,742	$1,192,212	$1,085,462	$3,463,416	$1,560,756	$1,507,305	$1,408,536	$4,476,598	$1,635,888	$1,550,523	$1,498,876	$4,685,287
Benefits												
Health benefits	$1,031,250	$1,072,500	$866,250	$2,970,000	$1,849,375	$1,746,250	$1,588,125	$5,183,750	$1,986,875	$1,876,875	$1,787,500	$5,651,250
Net benefits (health benefits − total costs)	−$154,492	−$119,712	−$219,212	−$493,416	$288,619	$238,945	$179,589	$707,152	$350,987	$326,352	$288,624	$965,963
Net benefits per patient	−$257	−$203	−$380	−$840	$481	$404	$308	$1,192	$585	$554	$496	$1,635
Total net benefits among population (N = 3,217)	−$276,111	−$217,211	−$407,397	−$900,718	$515,825	$432,818	$329,759	$1,278,402	$627,292	$594,157	$531,788	$1,753,237
Return on Nontargeted Outreach Interventions												
Total net benefits among population (N = 3,217)	$867,007	$809,764	$454,150	$2,130,921								

Notes: We extrapolated the calculation to the 3,217 patients eligible for randomization. Outreach costs = Number of call hours × Cost per hour. Screening costs = Number of ultrasounds completed × Cost per ultrasound + Number of CTs/MRIs completed × Average unit cost (CT/MRI). Treatment costs = Early detection probability × Average cost of treatment (if detected early). Opportunity costs = Annual HCC probability × Annual cost of advanced HCC. Health benefits = Quality-adjusted life gain × Financial value per quality-adjusted life year.

Table 6. Simulation Results Based on Patient-Level Treatment Effect Estimates.

A: Recommended Allocation

Condition	Usual Care	Outreach Alone	Outreach with Patient Navigation	Sample Size
Period 1	0%	99.9%	.1%	1,800
Period 2	9.0%	74.0%	16.9%	1,772
Period 3	8.4%	66.9%	24.7%	1,743

B: Return on Patient-Level Targeted Outreach Interventions

	Usual Care			Outreach Alone			Outreach with Patient Navigation		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
Cost and Benefit by Intervention Type and Across Period									
Sample size after reallocation	0	160	147	1,799	1,312	1,166	1	300	430
Net benefits per patient	−$257	−$203	−$380	$481	$404	$308	$585	$554	$496
Total net benefits among population (N = 3,217)	$0	−$58,838	−$103,076	$1,546,617	$961,383	$661,788	$1,045	$301,773	$393,578

Return on Patient-Level Targeted Outreach Interventions

	Period 1	Period 2	Period 3	Total (Improvement)
Total net benefits among population (N = 3,217)	$1,547,662	$1,204,318	$952,290	$3,704,270 (74%)

Return on Patient-Level Targeted Cancer Outreach Interventions: A Simulation

Thus far, the calculation of return on cancer interventions has been based on the random assignment of patients and on patients remaining in the same condition over three periods. However, (1) there is heterogeneity in patient-level treatment effects of outreach-alone intervention and in those of outreach-with-patient-navigation intervention, (2) not all patient-level treatment effect estimates are statistically greater than 0, (3) treatment effect heterogeneity varies across periods, and (4) the net return on outreach interventions varies across intervention types and over time. As such, given each patient's characteristics in a particular period, outreach with patient navigation is unlikely to be uniformly more effective than outreach alone. This poses two questions: (1) Given each patient's observed characteristics, which intervention type is most suitable for each patient? and (2) For the same patient, does the most suitable intervention vary across periods? Accordingly, we conduct a simulation that assigns each patient to the most suitable condition in each period based on two types of allocation schemes: (1) predicted treatment effect and (2) predicted net return (see detailed procedure in Web Appendix G).

Recommended allocation based on predicted treatment effect. Conceptually, in each period, given each patient's profile, we compare each patient's treatment effect estimate in their corresponding condition (e.g., condition 2) with their simulated treatment effect estimate in the counterfactual condition (e.g., condition 3). Then we assign this patient to the best-suited intervention that generates a significantly higher treatment effect estimate (e.g., condition 3, $p < .05$). If none of these estimates is significantly larger than 0, we assign this patient to condition 1.

As we show in Table 6, Panel A, the recommended allocation for each period is as follows:

1. Period 1: 0%, 99.9%, and .1% of patients are assigned to the no-outreach, outreach-alone, and outreach-with-patient-navigation conditions, respectively.
2. Period 2: 9.0%, 74.0%, and 16.9% of patients in each condition, respectively.
3. Period 3: 8.4%, 66.9%, and 24.7% of patients in each condition, respectively.

There are four noteworthy takeaways from this recommendation. First, the recommended split deviates from the original allocation based on the randomized controlled trial (1:1:1), suggesting that targeting induces asymmetric allocation of patients to different conditions. Second, there is a fraction of patients who stay in the baseline condition in Periods 2 and 3. For these patients, neither of the interventions is more effective than the baseline. Third, we reallocate most patients to the outreach-alone condition, suggesting that health care institutions can achieve the same level of effectiveness by aligning only moderate outreach efforts with these patients. Fourth, over

Table 7. Simulation Results Based on Patient-Level Estimated Return.

A: Recommended Allocation

Condition	Usual Care	Outreach Alone	Outreach with Patient Navigation	Sample Size
Period 1	0%	87.2%	12.8%	1,800
Period 2	0%	79.3%	20.7%	1,772
Period 3	0%	68.7%	31.3%	1,743

B: Return on Patient-Level Targeted Outreach Interventions

	Usual Care			Outreach Alone			Outreach with Patient Navigation		
	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3	Period 1	Period 2	Period 3
Cost and Benefit by Intervention Type and Across Period									
Sample size after reallocation	0	0	0	1,570	1,406	1,198	230	366	545
Net benefits per patient	−$257	−$203	−$380	$481	$404	$308	$585	$554	$496
Total net benefits among population (N = 3,217)	$0	$0	$0	$1,349,743	$1,030,263	$679,950	$240,462	$368,163	$498,837

Return on Patient-Level Targeted Outreach Interventions

	Period 1	Period 2	Period 3	Total (Improvement)
Total net benefits among population (N = 3,217)	$1,590,205	$1,398,426	$1,178,788	$4,167,419 (96%)

time, the outreach-with-patient-navigation condition seems to be more effective, given the higher allocation to this condition (.1% in Period 1, 16.9% in Period 2, and 24.7% in Period 3).

Table 6, Panel B, shows the return on patient-level targeted outreach interventions. When extrapolated to the 3,217 patients eligible for randomization from the hospital's patient database, patient-level targeted outreach program across conditions generates a gain of $1,547,662, $1,204,318, and $952,290 in Periods 1, 2, and 3, respectively. The total net return on patient-level targeted outreach program is $3,704,270, or 74% higher than that on nontargeted outreach program based on the random assignment ($3,704,270– $2,130,921 = $1,573,349).

Recommended allocation based on predicted net return. What if the health care system aims to maximize the overall return derived from assigning each patient to the most suitable intervention? As such, we can assign each patient to the intervention that gives the highest predicted net return based on Equation 6 rather than only the highest predicted treatment effect. Specifically, in each period, we compare each patient's predicted net return in the corresponding condition (e.g., condition 2) with their simulated patient's estimated net return in the counterfactual condition (e.g., condition 3). Then we assign the patient to the best-suited intervention that generates a significantly higher net return (e.g., condition 3, $p < .05$). If none of these estimates is significantly larger than 0, we assign this patient to condition 2 because the net return in the baseline condition is negative across all three periods (recall Table 5).

As we show in Table 7, Panel A, the recommended allocation for each period is:

1. Period 1: 0%, 87.2%, and 12.8% of patients are assigned to the no-outreach, outreach-alone, and outreach-with-patient-navigation conditions respectively.
2. Period 2: 0%, 79.3%, and 20.7% of patients in each condition, respectively.
3. Period 3: 0%, 68.7%, and 31.3% of patients in each condition, respectively.

There are two takeaways from this recommendation. First, no patient stays in the baseline condition under this allocation scheme, reflecting the goal of maximizing overall return. Second, while we still reallocate most patients to the outreach-alone condition, the outreach-with-patient-navigation intervention seems to be even more effective over time given the higher allocation to this condition than the previous allocation (e.g., 31.3% vs. 24.7% in Period 3).

Table 7, Panel B, shows the return on patient-level targeted outreach interventions. When extrapolated to the 3,217 patients eligible for randomization from the hospital's patient database, the total net return on the patient-level targeted outreach program is $4,167,419, or 96% higher than that on the nontargeted outreach program based on the random assignment ($4,167,419 − $2,130,921 = $2,036,498). In summary, patient-level targeted outreach interventions improve the payoffs to the health care system by 74%−96%, or $1.6 million to $2 million.

Discussion. The difference in allocation based on predicted treatment effect versus predicted net return shows the versatility of our approach in providing practical guidance to medical professionals and policy makers. The nature and magnitude of benefits can shift based on the goals that a health care institution sets for itself. Using this approach, an organization can set its strategic goals to maximize the benefits from personalized outreach. These results also confirm that the cumulated benefits from repeated and upgraded health education through outreach with patient navigation can be enhanced using individually tailored outreach over time.

Discussion

Relying only on the main-effects analysis, scholars might conclude that the outreach with patient navigation and outreach alone are equally effective. However, our application of causal forests uncovers patient heterogeneity in outreach effectiveness and leads to different conclusions and important practical implications. Specifically, patients with different characteristics respond very differently to each intervention. For example, patients who are more responsive to outreach alone or outreach with patient navigation tend to be female, be part of minority populations, be in better health status, be covered by medical assistance, have closer proximity to clinics, and reside in a populated neighborhood. Patients who are more responsive to outreach alone tend to be younger, have faster commutes, and reside in neighborhoods with more public insurance coverage. Patients responsive to outreach with patient navigation tend to be older and reside in a higher-income neighborhood. Over time, the outreach-with-patient-navigation intervention becomes more effective for an increased proportion of patients. As such, we illustrate time-varying heterogeneity in the outreach effectiveness.

A cost–benefit analysis shows that the baseline condition results in a net loss of $840 per patient, whereas outreach alone (outreach with patient navigation) generates a gain of $1,192 ($1,635) per patient. When extrapolated to the 3,217 eligible patients, the total net gain of the nontargeted cancer outreach program across conditions is $2,130,921, which implies that outreach marketing provides a substantial positive payoff to the health care system. Our simulation shows that targeted outreach interventions can enhance this return by 74%−96%.

Research Implications

For the marketing discipline, this article provides a framework for better understanding and analyzing sufficiently powered field experiments that are based on random assignment of heterogeneous customers to different treatments. Instead of focusing only on the main effects of the treatment or a subset of individual-level covariates, causal forests flexibly predict personalized treatment effects based on high-dimensional, nonlinear functions of those covariates. Such an approach also obviates the need for choosing several one-way interactions a priori to test for heterogeneity or searching over many interactions for particularly responsive subgroups. Accordingly, this article provides a methodological solution to the field's concern of external validity. Many empirical findings are typically much less generalizable than we imagine, because researchers lack a process and corresponding insights to identify moderators (i.e., the interaction of treatment and unmodeled/unmanipulated background factors) (Cook and Campbell 1979; Lynch 1982).

For the emerging discipline of personalized health care, we show that causal forests can identify particularly responsive subgroups without the need for a larger number of experimental conditions. While modern health care has implemented personalized medicine using genetic information, most health care outreach and educational programs still rely on untailored communications. Practitioners who manage these programs should recognize that the use of a large number of patient characteristics can substantially improve the outreach responsiveness through a tailored approach.

Our research also responds to a recent call for boundary-breaking marketing-relevant research (MacInnis et al. 2020) in several ways. First, our covariates are motivated by "real-world phenomena, rather than the constructs and theories in the marketing" (p. 11). Our findings that treatment effects vary across covariates not only engage "academics in other disciplines" (p. 1) but also offer important implications to the extant literature and theory going forward. Second, our findings have "life and death implications" (p. 9)—they help detect liver cancer at early stages. Third, our covariates, such as ethnicity, language, insurance coverage, and neighborhood socioeconomic status, elucidate how outreach effectiveness may vary among "understudied consumers such as minorities, privileged or impoverished classes, and marginalized consumers (e.g., special needs populations)" (p. 5).

Patient-Centric Health Care Marketing Implications

We urge hospitals and medical centers with outreach programs to leverage patient information to improve the effectiveness of outreach investments. Hospitals and health care practitioners should realize that a "one-size-fits-all" outreach program is neither effective nor economic. The use of machine learning can power data-driven patient-centric outreach programs that are also dynamically adaptive. Practitioners should consider both cross-sectional and temporal adaptation of outreach programs to maximize the benefit of health care interventions.

We urge policy makers in the federal, state, and local health departments, American Hospital Association, American Cancer Society, and American Liver Foundation to financially support personalized outreach programs. More hospitals should reach out to the underrepresented populations as they are more responsive to outreach messages; however, this requires additional resources and training. Incentivizing hospitals to reach out to patients with varying personal, clinical, structural, and socioeconomic backgrounds can also be effective. They should also engage a multidisciplinary group from health care,

marketing, computer science, and other disciplines to fund an accumulation of comprehensive databases to facilitate even better targeting of patients to improve outcomes.

Limitations and Future Research

First, because patients have different barriers to screening, future research should test the effectiveness of different barrier-reduction strategies by analyzing the nature of communication between patients and the staff with the use of call recordings. Second, our study focuses on the endpoint outcome: screening completion. Future research could apply the notion of customer journey to disentangle which parts of the intervention (e.g., barrier discussion during an outreach call vs. reminder calls) are more effective at not only increasing completion but also reducing no-show rates or time to response, further enhancing the return. Third, although we track individual patients, outreach designed to serve an individual may have influenced other members of the household. Future research could study possible spillover effects of outreach interventions.

References

American Cancer Society (2019), "Liver Cancer Survival Rates," (accessed February 25, 2020), https://www.cancer.org/cancer/liver-cancer/detection-diagnosis-staging/survival-rates.html.

Andersson, Karin L., Joshua A. Salomon, Sue J. Goldie, and Raymond T. Chung (2008), "Cost Effectiveness of Alternative Surveillance Strategies for Hepatocellular Carcinoma in Patients with Cirrhosis," *Clinical Gastroenterology and Hepatology*, 6 (12), 1418–24.

Arndt, Jamie, Clay Routledge, and Jamie L. Goldenberg (2006), "Predicting Proximal Health Responses to Reminders of Death: The Influence of Coping Style and Health Optimism," *Psychology and Health*, 21 (5), 593–614.

Aro, Arja R., Harry J. de Koning, Pilvikki Absetz, and Marjut Schreck (2001), "Two Distinct Groups of Non-Attenders in an Organized Mammography Screening Program," *Breast Cancer Research and Treatment*, 70 (2), 145–53.

Ascarza, Eva (2018), "Retention Futility: Targeting High-Risk Customers Might Be Ineffective," *Journal of Marketing Research*, 55 (1), 80–98.

Athey, Susan and Guido Imbens (2016), "Recursive Partitioning for Heterogeneous Causal Effects," *Proceedings of the National Academy of Sciences*, 113 (27), 7353–60.

Australian Digital Health Agency (2018), "Safe, Seamless and Secure: Evolving Health and Care to Meet the Needs of Modern Australia," (accessed February 28, 2020), https://conversation.digitalhealth.gov.au/sites/default/files/adha-strategy-doc-2ndaug_0_1.pdf

Basch, Charles E., Randi L. Wolf, Corey H. Brouse, Celia Shmukler, Alfred Neugut, Lawrence T. DeCarlo, et al. (2006), "Telephone Outreach to Increase Colorectal Cancer Screening in an Urban Minority Population," *American Journal of Public Health*, 96 (12), 2246–53.

Basil, Michael D. and William J. Brown (1997), "Marketing Aids Prevention: The Differential Impact Hypothesis Versus Identification Effects," *Journal of Consumer Psychology*, 6 (4), 389–411.

Benoit, Ronald M., Henrik Grönberg, and Michael James Naslund (2001), "A Quantitative Analysis of the Costs and Benefits of Prostate Cancer Screening," *Prostate Cancer and Prostatic Diseases*, 4 (3), 138–45.

Blanchard, Simon J., Wayne S. DeSarbo, A.Selin Atalay, and Nukhet Harmancioglu (2012), "Identifying Consumer Heterogeneity in Unobserved Categories," *Marketing Letters*, 23 (1), 177–94.

Bolton, Lisa E., Americus Reed II, Kevin G. Volpp, and Katrina Armstrong (2008), "How Does Drug and Supplement Marketing Affect a Healthy Lifestyle?" *Journal of Consumer Research*, 34 (5), 713–26.

Borsky, Amanda, Chunliu Zhan, Therese Miller, Quyen Ngo-Metzger, Arlene S. Bierman, and David Meyers (2018), "Few Americans Receive All High-Priority, Appropriate Clinical Preventive Services," *Health Affairs*, 37 (6), 925–28.

Breiman, Leo (2001), "Random Forests," *Machine Learning*, 45 (1), 5–32.

Bruix, Jordi and Morris Sherman (2011), "Management of Hepatocellular Carcinoma: An Update," *Hepatology*, 53 (3), 1020–22.

Bühlmann, Peter and Bin Yu (2002), "Analyzing Bagging," *The Annals of Statistics*, 30 (4), 927–61.

Caplan, Lee S., Barbara L. Wells, and Suzanne Haynes (1992), "Breast Cancer Screening Among Older Racial/Ethnic Minorities and Whites: Barriers to Early Detection," *Journal of Gerontology*, 47, 101–10.

Carlson, Rae (1971), "Sex Differences in Ego Functioning: Exploratory Studies of Agency and Communion," *Journal of Consulting and Clinical Psychology*, 37 (2), 267–77.

Chubak, Jessica and Rebecca Hubbard (2016), "Defining and Measuring Adherence to Cancer Screening," *Journal of Medical Screening*, 23 (4), 179–85.

Cole, Catherine A., Gilles Laurent, Aimee Drolet, Jane Ebert, Angela Gutchess, Raphaëlle Lambert-Pandraud, et al. (2008), "Decision Making and Brand Choice by Older Consumers," *Marketing Letters*, 19 (3/4), 355–65.

Conger, Krista (2010), "Rhoads Named Director of Cancer Outreach Programs for Stanford Cancer Center, Cancer Prevention Institute of California," Stanford Medicine News Center (March 1), http://med.stanford.edu/news/all-news/2010/03/rhoads-named-director-of-cancer-outreach-programs-for-stanford-cancer-center-cancer-prevention-institute-of-california.html.

Conger, Rand D., Katherine J. Conger, Glen H. Elder Jr., Frederick O. Lorenz, Ronald L. Simons, and Les B. Whitbeck (1992), "A Family Process Model of Economic Hardship and Adjustment of Early Adolescent Boys," *Child Development*, 63 (3), 526–41.

Cook, Thomas D. and Donald T. Campbell (1979), *Quasi-Experimentation Design & Analysis Issues for Field Settings*. Boston: Houghton Mifflin Company.

Efron, Bradley (2014), "Estimation and Accuracy After Model Selection," *Journal of the American Statistical Association*, 109 (507), 991–1007.

Ferrante, Jeanne M., Ping-Hsin Chen, and Steve Kim (2008), "The Effect of Patient Navigation on Time to Diagnosis, Anxiety, and Satisfaction in Urban Minority Women with Abnormal Mammograms: A Randomized Controlled Trial," *Journal of Urban Health: Bulletin of the New York Academy of Medicine*, 85 (1), 114–24.

Figueroa, Jose F. and Ashish K. Jha (2018), "Approach for Achieving Effective Care for High-Need Patients," *JAMA Internal Medicine*, 178 (6), 845–46.

Goldman, Shira N., David T. Liss, Tiffany Brown, Ji Young Lee, David R. Buchanan, Kate Balsley, et al. (2015), "Comparative Effectiveness of Multifaceted Outreach to Initiate Colorectal Cancer Screening in Community Health Centers: A Randomized Controlled Trial," *Journal of General Internal Medicine*, 30 (8), 1178–84.

Goossens, Nicolas, Amit G. Singal, Lindsay Y. King, Karin L. Andersson, Bryan C. Fuchs, Cecilia Besa, et al. (2017), "Cost-Effectiveness of Risk Score–Stratified Hepatocellular Carcinoma Screening in Patients with Cirrhosis," *Clinical and Translational Gastroenterology*, 8 (6), e101.

Grossman, Christopher H., Joanne Brooker, Natasha Michael, and David Kissane (2018), "Death Anxiety Interventions in Patients with Advanced Cancer: A Systematic Review," *Palliative Medicine*, 32 (1), 172–84.

Guo, Tong, S. Sriram, and Puneet Manchanda (2019), "The Effect of Information Disclosure on Industry Payments to Physicians," working paper.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2009), *The Elements of Statistical Learning: Prediction, Inference, and Data Mining*. New York: Springer-Verlag.

Humiston, Sharon G., Nancy M. Bennett, Christine Long, Shirley Eberly, Lourdes Arvelo, Joseph Stankaitis, et al. (2011), "Increasing Inner-City Adult Influenza Vaccination Rates: A Randomized Controlled Trial," *Public Health Reports*, 126 (2_suppl), 39–47.

Humphrey, Linda L., Mark Deffebach, Miranda Pappas, Christina Baumann, Kathryn Artis, Jennifer Priest Mitchell, et al. (2013) "Screening for Lung Cancer With Low-Dose Computed Tomography: A Systematic Review to Update the U.S. Preventive Services Task Force Recommendation," *Annals of Internal Medicine*, 159 (6), 411–20.

Johns Hopkins Medicine (2019), "The Sidney Kimmel Comprehensive Cancer Center: Outreach and Education Programs in Your Community," (accessed February 25, 2020), https://www.hopkinsmedicine.org/kimmel_cancer_center/centers/community_outreach/cancer_disparities_center/outreach_education.html%20.

Kaldenberg, Dennis O., Harold F. Koenig, and Boris W. Becker (1994), "Mail Survey Response Rate Patterns in a Population of the Elderly: Does Response Deteriorate with Age?" *Public Opinion Quarterly*, 58 (1), 68–76.

Khullar, Dhruv and Dave A. Chokshi (2018), "Health, Income, & Poverty: Where We Are & What Could Help," *Health Affairs* (October 4), https://www.healthaffairs.org/do/10.1377/hpb20180817.901935/full/.

King, Abby C., W.Jack Rejeski, and David M. Buchner (1998), "Physical Activity Interventions Targeting Older Adults: A Critical Review and Recommendations," *American Journal of Preventive Medicine*, 15 (4), 316–33.

Klein, William M.P. and Jennifer L. Cerully (2007), "Health-Related Risk Perception and Decision-Making: Lessons from the Study of Motives in Social Psychology," *Social and Personality Psychology Compass*, 1 (1), 334–58.

Lasser, Karen E., Jennifer Murillo, Sandra Lisboa, A. Naomie Casimir, Lisa Valley-Shah, Karen M. Emmons, et al. (2011), "Colorectal Cancer Screening Among Ethnically Diverse, Low-Income Patients: A Randomized Controlled Trial," *Archives of Internal Medicine*, 171 (10), 906–12.

Lisjak, Monika and Angela Y. Lee (2014), "The Bright Side of Impulse: Depletion Heightens Self-Protective Behavior in the Face of Danger," *Journal of Consumer Research*, 41 (1), 55–70.

Luce, Mary Frances and Barbara E. Kahn (1999), "Avoidance or Vigilance? The Psychology of False-Positive Test Results," *Journal of Consumer Research*, 26 (3), 242–59.

Lynch, John G. (1982), "On the External Validity of Experiments in Consumer Research," *Journal of Consumer Research*, 9 (3), 225–39.

MacInnis, Deborah J., Vicki G. Morwitz, Simona Botti, Donna L. Hoffman, Robert V. Kozinets, Donald R. Lehmann, et al. (2020), "Creating Boundary-Breaking, Marketing-Relevant Consumer Research," *Journal of Marketing*, 84 (2), 1–23.

Marrero, Jorge A., Laura M. Kulik, Claude B. Sirlin, Andrew X. Zhu, Richard S. Finn, Michael M. Abecassis, et al. (2018), "Diagnosis, Staging, and Management of Hepatocellular Carcinoma: 2018 Practice Guidance by the American Association for the Study of Liver Diseases," *Hepatology*, 68 (2), 723–50.

McCarthy, Ona, Irrfan Ahamed, Firuza Kulaeva, Ravshan Tokhirov, Salokhiddin Saibov, Marieka Vandewiele, et al. (2018), "A Randomized Controlled Trial of an Intervention Delivered by Mobile Phone App Instant Messaging to Increase the Acceptability of Effective Contraception Among Young People in Tajikistan," *Reproductive Health*, 15 (28), 1–14.

McMahon, Pamela M., Chung Yin Kong, Bruce E. Johnson, Milton C. Weinstein, Jane C. Weeks, Karen M. Kuntz, et al. (2008), "Estimating Long-Term Effectiveness of Lung Cancer Screening in the Mayo CT Screening Study," *Radiology*, 248 (1), 278–87.

Menon, Geeta, Lauren G. Block, and Suresh Ramanathan (2002), "We're at as Much Risk as We Are Led to Believe: Effects of

Message Cues on Judgments of Health Risk," *Journal of Consumer Research*, 28 (4), 533–49.

Moorman, Christine and Erika Matulich (1993), "A Model of Consumers' Preventive Health Behaviors: The Role of Health Motivation and Health Ability," *Journal of Consumer Research*, 20 (2), 208–28.

National Cancer Institute (2018), "Cancer Statistics," (accessed February 25, 2020), https://www.cancer.gov/about-cancer/understanding/statistics.

National Comprehensive Cancer Network (2018), "Hepatobiliary Cancers," (accessed February 26, 2020), https://www.nccn.org/patients/guidelines/content/PDF/hepatobiliary-patient.pdf.

Nyström, L., S. Wall, L.E. Rutqvist, A. Lindgren, M. Lindqvist, S. Rydén, et al. (1993), "Breast Cancer Screening with Mammography: Overview of Swedish Randomised Trials," *The Lancet*, 341 (8851), 973–78.

Oeffinger, Kevin C., Elizabeth T.H. Fontham, Ruth Etzioni, Abbe Herzig, James S. Michaelson, Ya-Chen Tina Shih, et al. (2015), "Breast Cancer Screening for Women at Average Risk: 2015 Guideline Update from the American Cancer Society," *JAMA*, 314 (15), 1599–1614.

Provincial Health Services Authority (2011), "Towards Reducing Health Iniquities: A Health System Approach to Chronic Disease Prevention" research report (accessed February 28, 2020), http://www.bccdc.ca/pop-public-health/Documents/TowardsReducingHealthInequitiesFinalDiscussionPape.pdf.

Rosenbaum, Paul R. and Donald B. Rubin (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," *Biometrika*, 70 (1), 41–55.

Rubin, Donald B. (1974), "Estimating Causal Effects of Treatments in Randomized and Nonrandomized Studies," *Journal of Educational Psychology*, 66 (5), 688–701.

Santi, Valentina, Franco Trevisani, Annagiulia Gramenzi, Alice Grignaschi, Federica Mirici-Cappa, Paolo Del Poggio, et al. (2010), "Semiannual Surveillance Is Superior to Annual Surveillance for the Detection of Early Hepatocellular Carcinoma and Patient Survival," *Journal of Hepatology*, 53 (2), 291–97.

Singal, Amit G., Anjana Pillai, and Jasmin Tiro (2014), "Early Detection, Curative Treatment, and Survival Rates for Hepatocellular Carcinoma Surveillance in Patients with Cirrhosis: A Meta-Analysis," *PLOS Medicine*, 11 (4), e1001624.

Singal, Amit G., Jasmin A. Tiro, Caitlin C. Murphy, Jorge A. Marrero, Katharine McCallister, Hannah Fullington, et al. (2019), "Mailed Outreach Invitations Significantly Improve HCC Surveillance Rates in Patients with Cirrhosis: A Randomized Clinical Trial," *Hepatology*, 69 (1), 121–30.

Singal, Amit G., Adam Yopp, Celette S. Skinner, Milton Packer, William M. Lee, and Jasmin A. Tiro (2012), "Utilization of Hepatocellular Carcinoma Surveillance Among American Patients: A Systematic Review," *Journal of General Internal Medicine*, 27 (7), 861–67.

Skolnik, Neil S. (2011), *Electronic Medical Records*. Totowa, NJ: Humana Press.

Sun, Chenshuo, Zijun (June) Shi, Xiao Liu, Anindya Ghose, Xueying Li, and Feiyu Xiong (2019), "The Effect of Voice AI on Consumer Purchase and Search Behavior," working paper.

Szczepura, A. (2005), "Access to Health Care for Ethnic Minority Populations," *Postgraduate Medical Journal*, 81 (953), 141–47.

Trinchet, Jean-Claude, Cendrine Chaffaut, Valérie Bourcier, Françoise Degos, Jean Henrion, Hélène Fontaine, et al. (2011), "Ultrasonographic Surveillance of Hepatocellular Carcinoma in Cirrhosis: A Randomized Trial Comparing 3- and 6-Month Periodicities," *Hepatology*, 54 (6), 1987–97.

Trudeau, Kimberlee J., Sharon Danoff-Burg, Tracey A. Revenson, and Stephen A. Paget (2003), "Agency and Communion in People with Rheumatoid Arthritis," *Sex Roles*, 49 (7/8), 303–31.

Tzartzeva, Kristina, Joseph Obi, Nicole E. Rich, Neehar D. Parikh, Jorge A. Marrero, Adam Yopp, et al. (2018), "Surveillance Imaging and Alpha Fetoprotein for Early Detection of Hepatocellular Carcinoma in Patients with Cirrhosis: A Meta-Analysis," *Gastroenterology*, 154 (6), 1706–18.

Wager, Stefan and Susan Athey (2018), "Estimation and Inference of Heterogeneous Treatment Effects Using Random Forests," *Journal of the American Statistical Association*, 113 (523), 1228–42.

Wager, Stefan, Trevor Hastie, and Bradley Efron (2014), "Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife," *Journal of Machine Learning Research*, 15 (1), 1625–51.

Wang, Wenbo, Hean Tat Keh, and Lisa E. Bolton (2010), "Lay Theories of Medicine and a Healthy Lifestyle," *Journal of Consumer Research*, 37 (1), 80–97.

Wetherell, Julie Loebach, Andrew J. Petkus, Steven R. Thorp, Murray B. Stein, Denise A. Chavira, Laura Campbell-Sills, et al. (2013), "Age Differences in Treatment Response to a Collaborative Care Intervention for Anxiety Disorders," *British Journal of Psychiatry*, 203 (1), 65–72.

Zeithaml, Valarie A., Bernard J. Jaworski, Ajay K. Kohli, Kapil R. Tuli, Wolfgang Ulaga, and Gerald Zaltman (2020), "A Theories-in-Use Approach to Building Marketing Theory," *Journal of Marketing*, 84 (1), 32–51.

Zhang, Bo-Heng, Bing-Hui Yang, and Zhao-You Tang (2004), "Randomized Controlled Trial of Screening for Hepatocellular Carcinoma," *Journal of Cancer Research and Clinical Oncology*, 130 (7), 417–22.