Working Paper

Effects of expanding health screening on treatment – What should we expect? What can we learn?

Rebecca Myerson*, Darius Lakdawalla, Lisandro D. Colantonio,

Monika Safford, David Meltzer

April 26, 2018

Abstract

Screening interventions can produce very different treatment outcomes, depending on the reasons why patients had been unscreened in the first place. Economists have paid scant attention to these complexities and their implications for evaluating screening programs. In this paper, we propose a simple economic framework to guide policy-makers and analysts in designing and evaluating the impact of screening on treatment uptake. We apply these insights to several salient empirical examples that illustrate the different kinds of effects screening programs might produce. Our empirical examples focus on contexts relevant to the top cause of death in the United States, heart disease. We find that currently undiagnosed patients differ from currently diagnosed patients in important ways, leading to lower predicted uptake of recommended treatment if these patients were diagnosed. Additionally, changes in the composition of diagnosed patients can produce misleading conclusions during policy analysis, such as spurious reductions in measured health system performance as screening expands.

Keywords: diagnosis, screening, self-selection, policy analysis, quality measurement. JEL Codes: I10, I12, I14, I18.

I Introduction

Many people – particularly those in vulnerable groups – suffer from undiagnosed conditions that result in missed opportunities to improve health. Diabetes, high cholesterol, and hypertension are chief contributors to avoidable premature mortality in the United States: despite available treatment, many patients with these conditions are undiagnosed or untreated (Cowie et al., 2009; Global Burden of Disease Collaborators, 2013; McDonald et al., 2009; Olives et al., 2013; Patel et al., 2015; Zweifler et al., 2011). Increasing access to screening for these chronic conditions has been a focus of both public-sector and private-sector efforts in recent years.[1]

Improving access to screening will at least weakly increase treatment of chronic conditions, because detection leads to treatment. However, the magnitude of this effect varies, because additional screening might diagnose patients with lower uptake of medical treatment after diagnosis. These gaps in treatment would shape the total costs and health benefits of policies and programs that subsidize screening, and affect physician practice after access to screening is expanded.

The key message of this paper is that screening policy can produce very different treatment and health outcomes, depending on the reasons why patients had remained unscreened in the first place. There are two main reasons that patients diagnosed after a screening expansion may have different treatment rates and outcomes from previously diagnosed patients. First, there is a clinical channel: additional screening may "over-diagnose" and result in treatment that does little to improve long-term health outcomes. For instance, epidemiologists have long observed that increases in cancer screening may fail to improve health outcomes at all, if it instead picks up small, slow-growing tumors that would never have harmed a patient's health (Ahn et al., 2014; Bleyer and Welch, 2012; Loeb et al., 2014).

---

[1]For example, Medicare added a free "Welcome to Medicare" visit for new enrollees in which screening needs are discussed and addressed, and the Affordable Care Act required health insurance plans to offer preventive care, including free screening for diabetes, high cholesterol, and hypertension to people at high risk. In the private sector, pharmacy chains such as CVS, Walgreens, and stores such as Ralph's and Sam's Club now offer screening for diabetes, high cholesterol, and hypertension in convenient retail locations.

Thus, screening programs may be reaching patients for whom the benefit of treatment is lower than average or even zero.

Second, there is an economic channel. Some patients' biomarkers are rarely assessed because they face high barriers to health care.[2] These barriers to care could include out-of-pocket costs, or non-pecuniary costs such as distance to a physician, language barriers, or psychological costs (Carpenter, 2010; Hyman et al., 1994; Lange, 2011; Kenkel, 1994; Manning et al., 1987; Musa et al., 2009). These same barriers could then translate to lower treatment rates for these patients after diagnosis, despite clinical benefit.

Economists have not studied the implications of the differences between currently diagnosed and undiagnosed patients for evaluating screening interventions. To fill this gap, we propose a simple economic framework to guide policy-makers and analysts in designing and evaluating the impact of screening policy on uptake of relevant treatment. We then apply the insights from this framework to several salient empirical examples that illustrate the different kinds of effects screening programs might produce.

The paper begins by presenting an economic framework of screening and treatment. Our framework is very parsimonious and requires two key assumptions: demand for screening is downward sloping, and screening provides the option of accessing treatment. As screening expands, newly diagnosed patients are expected to have higher ex ante net cost of treatment, i.e., higher cost or lower benefit to treatment, than previously diagnosed patients. Accordingly, these patients are predicted to have lower treatment uptake.

Our empirical exercises are tightly related to this theoretical framework and focus on heart disease risk factors which are commonly undiagnosed and untreated: diabetes, hypertension, and high cholesterol. We first test our prediction that as more patients are diagnosed, the additional diagnosed patients are expected to have higher cost or lower benefit to treatment than previously diagnosed patients. Using data from the National Health

---

[2]For evidence on screening, see Hyman et al. (1994); Lostao et al. (2001); Oster et al. (2013); Wilson (2011). For a discussion of self-selection into treatment and related econometric approaches, see, e.g., Carneiro et al. (2010); Eisenhauer et al. (2010); Heckman (2010).

4

and Nutrition Examination Survey (NHANES), we show that people whose blood glucose and blood cholesterol were not recently assessed tend to show larger barriers to treatment and/or appear ex ante to be healthier, with lower overall heart disease risk. Subsequently, these same factors are associated with lower propensity to treat high cholesterol and diabetes after diagnosis.

To quantify the magnitude of this effect, we compare the patients who are most vs. least likely to have their biomarkers assessed regularly (i.e., top 20th percentile vs. bottom 20th percentile). We find that after diagnosis, patients who were least likely to have their biomarkers assessed for undiagnosed conditions were also 49 percentage points less likely to use doctor-recommended treatment for high cholesterol and 32 percentage points less likely to receive a foot and eye exam for diabetes. This analysis provides support for our hypothesis that currently undiagnosed patients differ from currently diagnosed patients in important ways, leading to lower uptake of recommended treatment if these patients become diagnosed.

Our next empirical analyses explore the practical implications of these changes in the composition of diagnosed patients for policy analysis. Commonly used health system performance metrics focus on treatment and control of conditions that are diagnosed (Center for Medicare and Medicaid Services, 2011, 2016a; National Committee on Quality Assurance, 2016; Song et al., 2011, 2014). Diagnosed conditions are usually used for quality metrics because the true prevalence of conditions is not observed, whereas diagnosis status is observed.

We use two national datasets to show that these diagnosis-based treatment metrics often decline as diagnosis rates rise. In national Medicare data, hospital referral regions with higher diagnostic intensity, as calculated by Finkelstein et al. (2017), show lower use of maintenance care such as eye exams and hba1c checks for patients with diabetes, as calculated in the Dartmouth atlas of health care quality (The Center for the Evaluative Clinical Sciences, Dartmouth Medical School, 2017; Fisher et al., 2008). A repeated cross-section analysis using the NHANES data yields similar findings: time periods with higher diagnosis rates show lower rates of treatment among diagnosed conditions. Together, these analyses suggest

that changes in the composition of diagnosed patients as diagnosis rates rise could mask the benefits of screening expansions, as captured by commonly used measures of health system performance.

This paper provides several novel insights. First, based on the patterns we uncovered, expanded screening as a stand-alone program is likely to be less cost-effective than previously anticipated due to low treatment uptake among marginally screened patients. To our knowledge, these effects are not currently accounted for in cost-effectiveness analyses that simulate the impact of screening expansions (The CDC Diabetes Cost-Effectiveness Study Group, 1998; Glümer et al., 2006; Hoerger et al., 2004; Kahn et al., 2010; Nathan and Herman, 2004; Wang et al., 2011). Accounting for these effects could change the coverage policies selected in health systems that make decisions based on cost-effectiveness analysis.

Second, screening program design must account for the reasons why patients are unscreened. Numerous screening expansions have shown lower effects than anticipated, including programs in national health systems and screening programs for underinsured women (Ahn et al., 2014; Kim et al., 2017; Lantz et al., 1997). By showing how the health effects of diagnosing additional patients depend on the reasons why patients were undiagnosed in the first place, our framework clarifies when low health effects are avoidable and how they can be avoided.

Third, our research contributes to the small but growing literature on the unintended effects of quality reporting (Casalino et al., 2007; Dranove et al., 2003; Harris et al., 2016; Karve et al., 2008). In multiple pay-for-performance systems such as Accountable Care Organizations, providers have financial incentives to maintain high treatment rates for diagnosed conditions as well as high screening rates (Center for Medicare and Medicaid Services, 2011, 2016a; National Committee on Quality Assurance, 2016; Song et al., 2011, 2014). However, our research suggests that expanding access to screening could carry a penalty by reducing other, treatment-related quality metrics. This would suggest reconsideration or reweighting of the metrics used in pay-for-performance systems, to avoid penalizing health systems that

expand screening in diverse patient populations.

The paper proceeds as follows. Section II compares this study with previous literature. Section III presents our conceptual framework. We assess predictions from the framework empirically in section IV. Section V demonstrates the implications of our findings for policy analysis and health system performance measurement. Section VI concludes.

II   Comparison With the Literature

Anticipated costs and benefits of health care can differ across individuals, influencing individuals' willingness to seek care (Egan and Philipson, 2014; Eisenhauer et al., 2010; Heckman, 2010). This premise underlies commonly used public health models such as the health belief model.[3] It follows that anticipated net benefits of particular health services can vary across individuals (Vanness and Mullahy, 2012). In certain cases, distributions of these individual-level net benefits can be estimated (Basu and Heckman, 2007; Carneiro et al., 2010; Eisenhauer et al., 2010). These distributions are useful because changes to out-of-pocket costs of health care will attract different patients to treatment, depending on their anticipated cost and benefit (Basu and Meltzer, 2007; Goldman and Philipson, 2007; Pauly and Blavin, 2008).

A number of recent papers use new econometric methods to estimate distributions of net benefits of specific health services. These papers typically focus on how patients choose between treatments for their conditions (i.e., the intensive margin) (Basu and Heckman, 2007; Basu and Manning, 2009; Basu, 2011, 2013; Huang et al., 2006; Meltzer and Huang, 2007; Sculpher, 2008). In contrast, our theoretical model considers how the anticipated net benefits of screening are distributed in the population. These net benefits govern which conditions are *not* treated (i.e., the extensive margin). We present a simple and intuitive

---

[3]See Glanz and Bishop (2010) for a review of commonly used health behavior models in the public health field. The health belief model includes perceived benefits and perceived barriers as a key construct, and these are the constructs that are most strongly predictive of behavior in empirical tests (Rosenstock et al., 1988; Carpenter, 2010).

7

framework that describes cost, benefit, and the demand for care; it would be straightforward to extend these insights to dynamic models of health investment, e.g., the Grossman (1972) health capital model.

Our study can also be situated in the literature on screening and the demand for information. In the economics literature, the demand for screening has been empirically related to patient perceptions of disease risk and treatment effectiveness. Demand for screening is low if no effective treatment yet exists, and becomes higher once treatment is available (Oster et al., 2013; Wilson, 2011). Research examining conditions with available treatment has found that people who know about their elevated risk for a condition have higher demand for screening (Lange, 2011). In the public health literature, research on the health behavior model shows that participants who anticipate higher risk of the condition are more motivated to take action, whereas those who face logistical barriers are less likely to take action; these same variables are found to be predictive of screening (Carpenter, 2010; Hyman et al., 1994; Lostao et al., 2001).[4] Research on cancer screening in the medical and public policy literature has shown that screening expansions attract patients with less severe conditions, resulting in concern about overdiagnosis (Ahn et al., 2014; Kadiyala and Strumpf, 2016; Loeb et al., 2014). Yet at the same time, patients not reached at all by screening policies are sicker (Kim and Lee, 2017). This underscores the point that factors other than clinical need can play an important role in determining who is screened.

In summary, multiple strands of literature across disparate fields have shown that costs or benefits of treatment empirically predict patients' uptake of screening. This self-selection process implies that the composition of diagnosed patients should change as access to screening changes. Yet, economists have not studied the implications for evaluating screening policy. Our paper addresses this gap in the literature.

---

[4]Stigma related to testing can also play a role, although this is less relevant in the disease contexts from which we draw our empirical examples (Godlonton and Thornton, 2012).

Figure 1: Demand for screening with a single supply curve
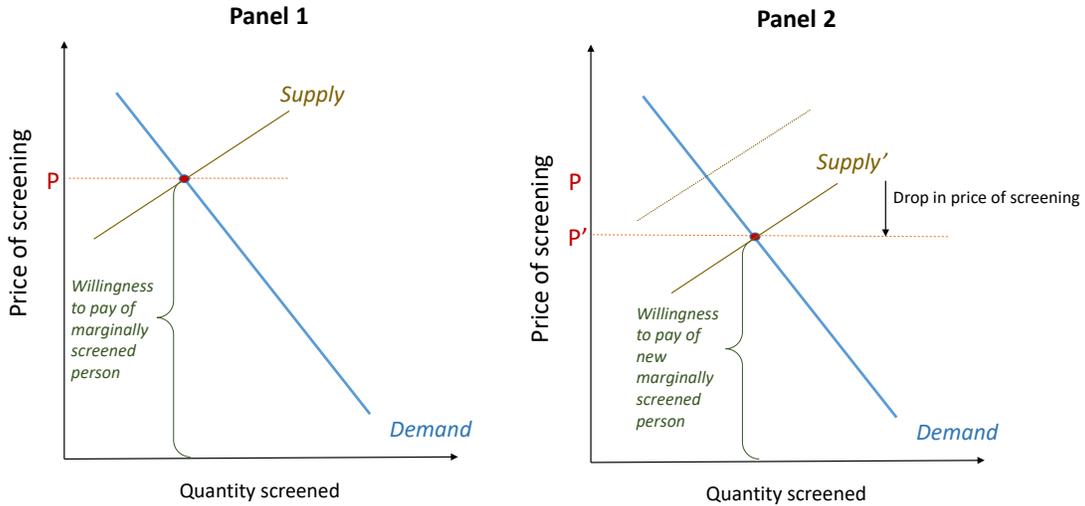
III Conceptual Framework

In this section, we present a simple economic framework of selection into screening. We employ two key assumptions: demand for screening is downward sloping, and screening provides the option of accessing medical treatment. This framework is then used to discuss the different kinds of effects screening programs might produce. Although our treatment of the theory is intuitive for clarity of presentation, similar points could be made using a formal model.

A Which patients are screened?

Suppose the value of screening differs across patients. In some cases, this variation occurs for reasons that are unobservable to patients or their physicians. For instance, some patients might be genetically predisposed to developing complications from a disease if untreated while others might be relatively immune to complications. Variation that is unobservable to the patient and her physician is less problematic for policy analysis because it will not

Figure 2: Demand for screening with multiple supply curves

Patients with low access
to care face higher price
of screening

Price of screening

Demand for screening

Quantity screened
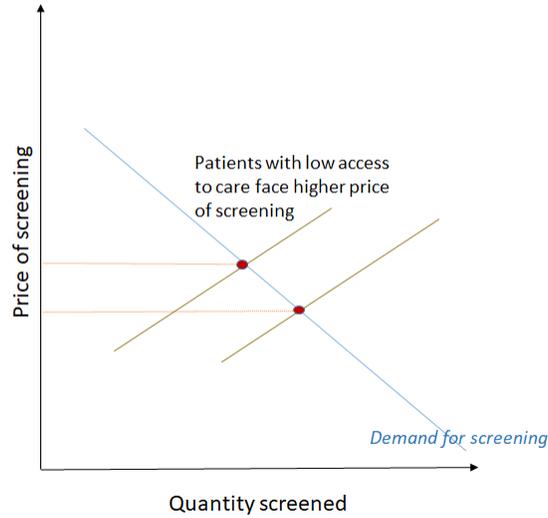
systematically affect a patient's propensity to be screened.

Greater complexity arises when this variation is observable to the patient or her physician. In this case, patients with high expected value from screening have higher willingness to pay to be screened. In economic terms this is a downward sloping demand curve, as in Figure 1. The vertical axis here indicates the price of screening and horizontal axis denotes quantity of people screened.

If all patients face the same price of screening, say price P in the first panel of Figure 1, then the patients with willingness to pay greater than or equal to P are screened, and patients with a lower willingness to pay are not screened. In the second panel of Figure 1, the price of screening shifts down to P'. In this case, many more patients are screened including those with lower willingness to pay, reflecting lower expected value.

In practice, not all patients face the same price of screening. Some patients face high out-of-pocket costs, live farther from a facility, face language barriers, or live in an area with low health care supply. Figure 2 shows the demand curve for screening again but with two supply curves added, reflecting the fact that different patients can face different pecuniary or non-pecuniary costs of screening. Based on the points of intersection between the supply and demand curves, fewer patients are screened in the group facing higher prices for screening.

Figure 3: Screening decision tree
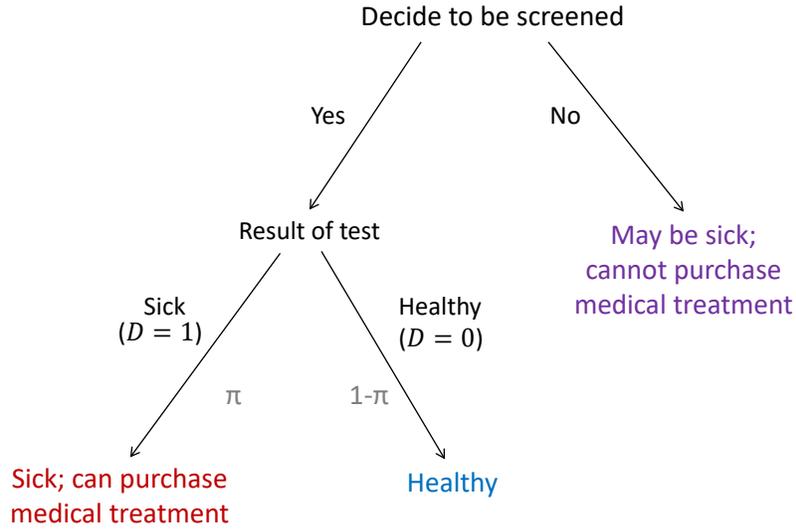
For each group, the logic of Figure 1 still applies: if one supply curve were to shift so as to lower the price of screening in one group, the marginally screened patients in that group would expect less value from screening than the previously screened patients.

B Where do patients sit on the demand curve?

We next consider the determinants of demand for screening. Why might some patients (in consultation with their doctors) expect more value from screening than other patients? This topic has been analyzed in the public health and economics literature (Boozer and Philipson, 2000; Hyman et al., 1994; Lostao et al., 2001; Oster et al., 2013; Wilson, 2011). We provide a brief discussion here with the aid of a diagram, Figure 3.

Screening is valuable to patients as an input to improved health and well-being (Glanz and Bishop, 2010; Carpenter, 2010). Figure 3 depicts one way screening can improve health, namely, by providing patients with the option to pursue medical treatments that are only available after diagnosis. (Hereafter, we will use the term "prescription-only medical treat-

ment" or just "treatment.") This is represented by the left-most branch of Figure 3. Ex post, this option is not used by patients with a negative screening result, as shown in the middle branch of Figure 3. Patients who have not been recently screened do not know their disease state and cannot access treatment, as shown in the right branch of Figure 3. The key take-away from this figure is that the value of screening is entwined with the value of treatment, because screening is a gateway to treatment.

Which patients expect little value from the option to pursue treatment? Two general predictions are proposed. First, patients who know they would have difficulties accessing or affording treatment – including pecuniary costs as well as non-pecuniary costs such as language barriers or distance to a facility – should find the option to pursue treatment less valuable, all else equal. Second, patients who think they are unlikely to benefit from treatment – either because they are unlikely to have the condition, or because the treatment is unlikely to work – should find screening less valuable, all else equal (Kim and Lee, 2017). Accordingly, demand for screening is low for conditions which lack effective treatments, such as or HIV before the development of highly active anti-retroviral therapy, or Huntington's disease (Oster et al., 2013; Wilson, 2011). In summary, the option to be treated is less valuable to patients who anticipate high pecuniary or non-pecuniary costs of treatment and/or little clinical benefit.

C What can we expect as screening expands?

Section III.B argued that a because a patient's expected costs and benefits of treatment contribute to the patient's expected value of screening, these factors also affect where that patient sits on the screening demand curve. Below, we argue that these same factors can produce declining treatment rates for newly diagnosed conditions as screening expands.

For simplicity, first consider the case where all patients face the same price of screening. In this case, we expect marginally screened patients, as screening expands, to have lower uptake of treatment after diagnosis. This point can be demonstrated by revisiting Figure 1. Suppose

that the price of screening starts very high and is slowly lowered due to a national screening program, screening mandates for insurance plans, or other interventions. When the price of screening is high, patients who anticipate high clinical benefits and low treatment costs may be willing to pay. As the price of screening declines, more patients are screened, including those who anticipate less clinical benefit or who would have difficulty affording treatment if diagnosed. Because of these higher costs or lower benefits to treatment, the newly diagnosed patients would treat their conditions at lower rates than previously diagnosed patients, on average – and the impact of screening on treatment would not be as large as one might otherwise expect.

Let's now consider the case where different patients face different prices for screening. Could high treatment rates be obtained by targeting a screening intervention to the less-screened group, since they are higher on the demand curve? Yes, if barriers to screening are uncorrelated with barriers to treatment or benefits to treatment. This might occur, for example, if a new screening technology is introduced in a national health system, and some regions have lower access to the technology than other regions for idiosyncratic reasons. Yet, such a situation would be highly unusual. In a far more empirically likely case, patients facing high barriers to screening (for example, patients who are uninsured or live in rural areas) will also face high barriers to treatment. In an insurance scheme where patient co-payments are set based on value, patients facing high prices should also show lower clinical benefits. Both factors reinforce our original prediction that currently undiagnosed patients differ from currently diagnosed patients in ways that predict lower uptake of treatment after diagnosis.

We assess these predictions empirically in section IV and find that currently undiagnosed patients indeed differ from currently diagnosed patients in important ways, leading to lower predicted uptake of recommended treatment. In section V, we explore the practical implications of these changes in the composition of diagnosed patients for policy analysis.

13

IV Intermediate Factors Linking Diagnosis and Treatment

As noted above, the patients who are tested for undiagnosed conditions should be the patients with a higher ex ante net benefit (i.e., benefit net of cost) of doing so. We hypothesized that some of the factors determining patients' ex ante cost and benefit of biomarker checks would also influence patients' take-up of treatment uptake after diagnosis.

In this section, we approach this topic empirically by comparing the factors associated with biomarker assessments with the factors associated with treatment after diagnosis for diabetes and high cholesterol. The goal of this analysis is to demonstrate that observable intermediate factors could attenuate the impact of screening expansions and quantify the size of this effect. The results are therefore informative for predicting the impact of policies that change access to biomarker assessment across a broad population.

A Data

We used data from the National Health and Nutrition Examination Survey (NHANES). NHANES is a nationally representative biomarker survey run by the Center for Disease Control and Prevention. Comparable data have been collected on a rolling basis from 1999-2014, and these are the data most commonly used to track awareness of chronic conditions over time on the national level (Centers for Disease Control and Prevention, 2014). Data on recent biomarker assessments relevant to diabetes are only available starting in 2005 and not fully available for years after 2014; we therefore used data from 2005-2014.

Diagnosed vs. undiagnosed conditions We coded participants as having a particular chronic condition if they reported prior diagnosis for the condition at the time of participation, with the appropriate exclusions for diagnosis during pregnancy, or if their biomarkers met standard definitions for the condition after taking their fasting status into account

(American Diabetes Association, 2014; Stone et al., 2014; James et al., 2014). Individuals were classified as undiagnosed for the condition if they met the biomarker definitions for a condition, but report no prior diagnosis for that condition. Table 4 in the Appendix includes details of each definition.

The relatively colloquial wording of the NHANES questions helps to mitigate concerns that patients may not report a prior diagnosis because they are unfamiliar with medical terms. Participants in the NHANES data were considered to be diagnosed with diabetes if they responded positively to the question, "Have you ever been told by a doctor that you have diabetes or sugar diabetes?" Participants were considered to be diagnosed with high cholesterol if they responded positively to the question, "Have you ever been told by a doctor or other health professional that your blood cholesterol level was high?"

Recent biomarker assessments relevant to undiagnosed conditions The NHANES data include information on recent biomarker assessments relevant to high cholesterol and diabetes. Biomarker assessments are tightly related to our topic of study, as the key step through which undiagnosed conditions become diagnosed. On the one hand, patients whose biomarkers are never assessed will never become diagnosed. On the other hand, physicians who observe biomarker assessment results indicative of previously undiagnosed hyperlipidemia or diabetes must ethically proceed according to national guidelines for diagnosis and treatment.[5]

Recent blood cholesterol assessment was captured using two questions in the NHANES: "Have you ever had your blood cholesterol checked?" and "About how long has it been since you last had your blood cholesterol checked? Has it been..." with the options "Less than a

---

[5]This point can be illustrated with multiple examples. For example, a physician may run a blood test to assess whether a patient who presents with symptoms of diabetes, such as polydipsia or polyuria, in fact has undiagnosed diabetes. Although this test may not have been scheduled in advance, the purpose of this test is to assess whether the patient has undiagnosed diabetes. As another example, a physician may conduct blood tests to track the effectiveness of treatment of diagnosed pre-diabetes that may or may not ultimately develop into diagnosed diabetes. Although one purpose of this test is to monitor the impacts of the treatment for pre-diabetes, ethically the physician must diagnose and treat diabetes if test results are indicative of diabetes.

year ago," "1 year but less than 2 years ago," "2 years but less than 5 years ago," or "5 years or more." We present results using the two-year look-back period, but findings are similar if other look-back periods are used.

Recent biomarker assessment related to diabetes was captured using the question, "Have you had a blood test for high blood sugar or diabetes within the past three years?" (In the NHANES instructions, interviewers coded patients with only urine tests as not having been tested for high blood sugar or diabetes.) The colloquial wording of this question helps to prevent false negative reports among patients unfamiliar with the word "diabetes". Nonetheless, some misclassification of biomarker assessment status or diagnosis status may remain, which would bias our findings toward the null (Black et al., 2000).

We created variables to capture biomarker assessment that would provide the opportunity for undiagnosed conditions to become diagnosed. Taking the example of high cholesterol, the biomarker assessment variable took the value 1 for patients who have never been diagnosed for high cholesterol and have recently had a cholesterol check, and took the value 0 for patients who had never been diagnosed for high cholesterol and have not recently had a cholesterol check. This variable was coded as missing for patients already diagnosed for high cholesterol. A similar variable was constructed for diabetes.

Uptake of recommended treatment after diagnosis We also created variables to capture the uptake of recommended treatment among patients who have already been diagnosed. Because pharmaceutical treatment is not recommended for all patients, we restricted the high cholesterol sample to only include patients who reported that their doctor had recommended taking medication for their high cholesterol. For diabetes, we focus on eye exams and foot exams, which are recommended for all patients with diabetes even when prescription medications are not. We analyzed each condition separately.

B   Analysis

Our statistical analysis proceeded as follows. First, we used correlation coefficients to demonstrate which factors are predictive of biomarker assessment. Next, we assessed whether these same factors could account for lower treatment rates among patients diagnosed after a policy change using a simulation exercise. To minimize assumptions, we used patients' real treatment outcomes. Our exercise therefore involved selectively dropping the treatment data of patients who had in fact been diagnosed, rather than imputing the potential treatment rates of undiagnosed patients.

For example, in one counterfactual, we only allowed patients in the bottom 20 percent of biomarker assessment propensity scores to become diagnosed, and all other diagnosed patients are dropped from the sample. In another counterfactual, all patients with propensity scores from the 20th to 40th percentile were allowed to be diagnosed and others were dropped. In each case, we calculate the average treatment rate of diagnosed conditions using patients' real data, including only patients who would have been diagnosed under such a scheme. We repeated this exercise for each 20-percentile bin to learn about the groups of patients who might become diagnosed under small vs. large changes in policy.

C   Results

The results indicate that, consistent with our hypotheses, patients who had higher barriers to care and who (ex ante) seemed healthier were less likely to be checked for undiagnosed conditions, and also less likely to be treated after diagnosis. Table 1 summarizes the sign of the statistically significant correlation coefficients found in the data. These correlations show that on average, patients who lacked recent blood tests relevant to undiagnosed conditions had larger barriers to care: these patients were more likely to lack health insurance, lack a usual source of care, and have low income. These same factors are, in turn, associated with lower likelihood of treating diagnosed conditions. Likewise, patients who appeared healthier - patients who were younger, had better self-reported health or lower Framingham risk score,

Table 1: Patient-level characteristics associated with biomarker checks for undiagnosed conditions are also associated with increased treatment of diagnosed conditions (NHANES data)

	Biomarkers Checked If Not Diagnosed	Treated If Diagnosed
Barriers to care		
No health insurance	↓	↓
No usual place for health care	↓	↓
Lower income	↓	↓
Benefits to care		
Poor self-reported health	↑	↑
Recently hospitalized overnight	↑	↑
Co-morbid condition	↑	↑
Framingham risk score	↑	↑
Age	↑	↑

This table shows that patient-level characteristics associated with biomarker checks are also associated with increased treatment after diagnosis in the NHANES data. The arrows represent the sign of statistically significant correlations between the variable listed in the column title and the variable listed in the row title. The sign and significance of the findings are the same for both diabetes and high cholesterol, two conditions with the requisite variables measured in the NHANES data.

fewer co-morbid conditions, or no recent hospitalizations - were less likely to receive blood tests relevant to any undiagnosed conditions and also less likely to treat their diagnosed conditions.

The correlation analyses therefore suggest that patient composition effects could shape the treatment rates for diagnosed patients as more patients become diagnosed. The simulation analyses confirm and quantify this phenomenon. For example, we can compare patients in the top 20th percentile vs. bottom 20th percentile in their likelihood for biomarker assessment. Such a comparison indicates that the patients who were least likely to have their biomarkers assessed for undiagnosed high cholesterol were also 49 percentage points less likely to use doctor-recommended treatment for high cholesterol after diagnosis. Findings were similar for diabetes: the patients who were least likely to have their biomarkers assessed for undiagnosed diabetes were also 32 percentage points less likely to receive an annual foot and eye exam for diabetes after diagnosis. See Table 2.

Table 2: Propensity to have biomarkers checked for undiagnosed conditions and treatment rates for diagnosed conditions (NHANES data)

High cholesterol	
Propensity to have biomarkers checked for undiagnosed high cholesterol	Take recommended high cholesterol medication if diagnosed with high cholesterol and medication is recommended
20th percentile or below	88%
20th percentile-40th percentile	79%
40th percentile-60th percentile	69%
60th percentile-80th percentile	48%
80th percentile and above	39%

Diabetes	
Propensity to have biomarkers checked for undiagnosed diabetes	Foot exam and eye exam in past year if diagnosed with diabetes
20th percentile or below	43%
20th percentile-40th percentile	35%
40th percentile-60th percentile	29%
60th percentile-80th percentile	19%
80th percentile and above	11%

This table shows that the patients who were less likely to have a blood test to check for undiagnosed diabetes or high cholesterol, based on their observed characteristics from Table 1, were also less likely to receive recommended treatment for these conditions after diagnosis.

The empirical findings in this section therefore support our hypothesis that the composition of diagnosed patients changes as diagnosis rates rise, in a way that can decrease treatment rates for diagnosed conditions.

V Implications for Measuring Health System Performance

The previous section provided evidence that the impact of screening expansions can be undermined by patient composition effects. Such a phenomenon could help to explain the numerous previous studies showing low impact of national cancer screening programs on treatment and health (Ahn et al., 2014; Kim and Lee, 2017; Lantz et al., 1997; Loeb et al., 2014).

This section considers the potential consequences of these patient composition effects for measurements of health system performance. A number of health system performance metrics focus on treatment and control of diagnosed conditions (Center for Medicare and Medicaid Services, 2011, 2016b; National Committee on Quality Assurance, 2016). This occurs because true prevalence of conditions is not measured in most datasets, whereas diagnosis status is frequently measured. However, we will argue that tracking the rate of treatment given diagnosis could lead to misleading conclusions during policy analysis.

Our first analysis focuses on treatment quality metrics from the Dartmouth atlas of health care's dataset on quality and effective care (The Center for the Evaluative Clinical Sciences, Dartmouth Medical School, 2017). These data provide a summary of quality of care metrics among Medicare beneficiaries by hospital referral region, including the fraction of patients with diagnosed diabetes who receive recommended medical care. We combined these data with data from Finkelstein et al. (2017) which summarize doctors' propensity to diagnose conditions ("diagnostic intensity") across hospital referral regions, after adjusting for the underlying health needs of the population. Table 3 shows the correlations between

Table 3: Bivariate relationship between diagnostic intensity and treatment of diagnosed diabetes (Dartmouth Atlas data)

| Outcome	Correlation with diagnostic intensity score
Average annual percent of Medicare enrollees with diabetes age 65-75 having eye examination	-0.22***
Average annual percent of Medicare enrollees with diabetes age 65-75 having hba1c test	-0.31***

* $p <0.1$, ** $p <0.05$, *** $p <0.01$

This table shows that hospital referral regions that have higher diagnostic intensity for reasons unrelated to the health of their population, as measured by Finkelstein et al. (2017), also have lower uptake of recommended treatment for diabetes as measured by the Dartmouth Atlas (The Center for the Evaluative Clinical Sciences, Dartmouth Medical School, 2017).

diagnostic intensity and appropriate treatment of diagnosed conditions. The results indicate that hospital referral regions with higher diagnostic intensity show lower rates of appropriate diabetes treatment for diagnosed conditions. In other words, diagnosis-based treatment metrics are low when diagnosis rates are high.

Similarly, national trends over time indicate that diagnosis-based treatment metrics decline as diagnosis rates rise. Figure 4 depicts nationally representative estimates of the following three quantities of interest, calculated from the NHANES data: (a) total prevalence on the population level, including undiagnosed conditions; (b) the fraction of people who truly have the condition who report being diagnosed, (c) and the fraction of people who are diagnosed for the condition who report taking medication to treat the condition. These data demonstrate that an increase in diagnosis of diabetes, hypertension, and high cholesterol in recent years coincided with a decrease in rates of treatment among diagnosed conditions.

Figure 4: Increased diagnosis of chronic conditions is associated with decreased rates of medical treatment for diagnosed conditions (NHANES data)

Source: Authors' analysis of NHANES data.
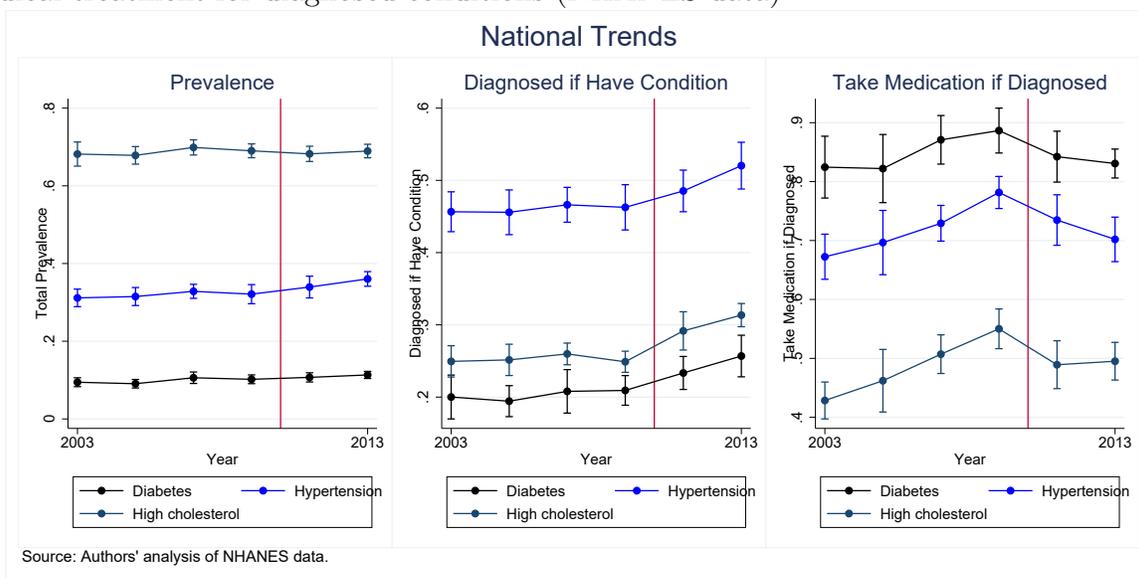
This figure uses repeated cross-sectional data from the NHANES study to demonstrate that an increase in diagnosis of diabetes, hypertension, and high cholesterol on the national level in recent years coincided with a fall in treatment rates among diagnosed conditions. The left panel depicts total prevalence, including undiagnosed conditions; the middle panel depicts the fraction of prevalent conditions that are diagnosed; and the right panel depicts the fraction of diagnosed conditions that are treated with medications.

VI Conclusion

The key message of this paper is that if anticipated costs and benefits of treatment determine which patients had remained undiagnosed prior to screening interventions, they should also shape the impact of screening policy. In locations with unequal access to care, such as the United States, barriers sometimes can play a more important role than clinical severity in determining which patients are screened and treated. Regardless, clinical benefits and barriers to care can both create a situation where treatment uptake diminishes as screening expands. This finding has a number of practical implications.

First, our framework and our findings have implications for policy analysis. Based on the patterns we uncovered, expanded screening as a stand-alone program is likely to be less cost-effective than previously anticipated due to low treatment uptake among marginally screened patients. To our knowledge, these effects are not currently accounted for in cost-effectiveness analyses that simulate the impact of screening expansions, and accounting for these effects could change the coverage policies selected in health systems that make decisions based on cost-effectiveness analysis. Additionally, when analyzing public policies that expand access to screening, we caution against using treatment of diagnosed conditions as an outcome metric. Analysts who use treatment of diagnosed conditions as an outcome metric in policy analysis could risk conflating changes in the composition of diagnosed patients with declines in health system performance.

Our findings also inform policy design in multiple ways. In pay-for-performance systems where providers have financial incentives to maintain high treatment rates for diagnosed conditions, such as Accountable Care Organizations, expanding access to screening could carry a penalty by reducing other quality metrics. This would suggest reconsideration or reweighting of the metrics used in pay-for-performance systems, to avoid penalizing health systems that expand screening in diverse patient populations. Finally, our analysis raises questions about how to best improve the uptake of needed health services among undiagnosed patients with low access to care.

Appendix table

A   Supplemental table

Table 4: Definitions used for diabetes, hypertension, and high cholesterol

Condition	Status	Definition
Diabetes	No condition	No self-reported diagnosis of diabetes and FPG<126 mg/dl or NFPG<200mg/dl
	Undiagnosed	No self-reported diagnosis of diabetes, but FPG>126 mg/dl or NFPG>200mg/dl
	Diagnosed	Self-reported diagnosis of diabetes (when non-pregnant for women)
Hypertension	No condition	No self-reported diagnosis, SBP<140mmHg, and DBP<90mmHg
	Undiagnosed	No self-reported diagnosis of hypertension, but SBP>140mmHg or DBP>90mmHg
	Diagnosed	Self-reported diagnosis of hypertension (when non-pregnant for women)
High cholesterol	No condition	No self-reported diagnosis, total cholesterol <200 mg/dl, LDL cholesterol<160 mg/dl, and HDL cholesterol>40 mg/dl
	Undiagnosed	No self-reported diagnosis, but total cholesterol >200 mg/dl, LDL cholesterol>160 mg/dl, or HDL cholesterol<40 mg/dl
	Diagnosed	Self-reported diagnosis

Note: FPG=fasting plasma glucose; NFPG=non-fasting plasma glucose; SBP=systolic blood pressure; DBP=diastolic blood pressure; HDL=high-density lipoprotein, LDL= low-density lipoprotein.

References

Agency for Healthcare Research and Quality (2013). 2012 National Healthcare Quality Report. Technical report, US Department of Health and Human Services, Rockville, MD.

Ahn, H. S., H. J. Kim, and H. G. Welch (2014). Korea's thyroid-cancer "epidemic" - screening and overdiagnosis. *New England Journal of Medicine 371*(19), 1765–1767. PMID: 25372084.

American Diabetes Association (2014, January). Standards of Medical Care in Diabetes–2014. *Diabetes Care 37 Suppl 1*(October 2013), S14–80.

Basu, A. (2011). Economics of individualization in comparative effectiveness research and a basis for a patient-centered health care. *Journal of Health Economics*.

Basu, A. (2013). Estimating person-centered treatment effects using instrumental variables: an application to evaluating prostate cancer. *Journal of Applied Econometrics*.

Basu, A. and J. Heckman (2007). Use of instrumental variables in the presence of heterogeneity and self selection: an application to treatments of breast cancer patients. *Health Economics 1157*(October), 1133–1157.

Basu, A. and W. Manning (2009). Issues for the next generation of health care cost analyses. *Medical Care 47*(7).

Basu, A. and D. Meltzer (2007). Value of information on preference heterogeneity and individualized care. *Medical Decision Making 27*(2), 112–27.

Black, D. A., M. C. Berger, and F. A. Scott (2000). Bounding parameter estimates with nonclassical measurement error. *Journal of the American Statistical Association 95*(451), 739–748.

Bleyer, A. and H. G. Welch (2012). Effect of three decades of screening mammography on breast-cancer incidence. *New England Journal of Medicine 367*(21), 1998–2005. PMID: 23171096.

Boozer, M. A. and T. J. Philipson (2000). The impact of public testing for human immun-

odeficiency virus. *The Journal of Human Resources 35*(3), 419–446.

Carneiro, P., J. Heckman, and E. Vytlacil (2010). Estimating marginal returns to education. *101*(October), 2754–2781.

Carpenter, C. J. (2010, December). A meta-analysis of the effectiveness of health belief model variables in predicting behavior. *Health Communication 25*(8), 661–9.

Casalino, L. P., A. Elster, A. Eisenberg, E. Lewis, J. Montgomery, and D. Ramos (2007, May). Will Pay-For-Performance And Quality Reporting Affect Health Care Disparities? *Health Affairs 26*(3), w405–w414.

CDC (2012). CDC - A Diabetes Report Card for the United States Quality of care in the 1990s - Publications - Diabetes DDT. *136*(8), 565–574.

Center for Medicare and Medicaid Services (2011). Guide to Quality Performance Scoring Methods for Accountable Care Organizations.

Center for Medicare and Medicaid Services (2016a). Publicly available ACO data and ACO performance data sources maintained by CMS. (March).

Center for Medicare and Medicaid Services (2016b). Quality Measures, Reporting and Performance Standards.

Centers for Disease Control and Prevention (1999-2014). National Health and Nutrition Examination Survey.

Cowie, C., K. Rust, and E. Ford (2009). Full accounting of diabetes and pre-diabetes in the US population in 1988-1994 and 2005-2006. *Diabetes . . . 32*(2), 0–7.

Dale, S. B., A. Ghosh, D. N. Peikes, T. J. Day, F. B. Yoon, E. F. Taylor, K. Swankoski, A. S. O'Malley, P. H. Conway, R. Rajkumar, M. J. Press, L. Sessums, and R. Brown (2016). Two-year costs and quality in the comprehensive primary care initiative. *New England Journal of Medicine 374*(24), 2345–2356. PMID: 27074035.

Dranove, D., D. Kessler, M. McClellan, and M. Satterthwaite (2003, June). Is More Information Better? The Effects of and quot;Report Cards" on Health Care Providers. *Journal of Political Economy 111*(3), 555–588.

Egan, M. and T. Philipson (2014, July). Health Care Adherence and Personalized Medicine. NBER Working Papers 20330, National Bureau of Economic Research, Inc.

Eisenhauer, P., J. Heckman, and E. Vytlacil (2010). Generalized Roy Model and the Cost-Benefit Analysis of Social Programs. *Policy Analysis*.

Finkelstein, A., M. Gentzkow, P. Hull, and H. Williams (2017). Adjusting risk adjustment - accounting for variation in diagnostic intensity. *New England Journal of Medicine 376*(7), 608–610. PMID: 28199802.

Fisher, A. E. S., D. C. Goodman, and A. Chandra (2008). Regional and Racial Variation in Health Care among Medicare Beneficiaries: A Brief Report of the Dartmouth Atlas Project. Technical report, The Dartmouth Institute for Health Policy and Clinical Practice.

Glanz, K. and D. Bishop (2010, January). The role of behavioral science theory in development and implementation of public health interventions. *Annual Review of Public Health 31*, 399–418.

Global Burden of Disease Collaborators (2013, August). The state of US health, 1990-2010: burden of diseases, injuries, and risk factors. *JAMA 310*(6), 591–608.

Glümer, C., M. Yuyun, S. Griffin, D. Farewell, D. Spiegelhalter, A. L. Kinmonth, and N. J. Wareham (2006). What determines the cost-effectiveness of diabetes screening? *Diabetologia 49*, 1536–1544.

Godlonton, S. and R. Thornton (2012). Peer effects in learning hiv results. *Journal of Development Economics 97*(1), 118 – 129.

Goldman, D. and T. Philipson (2007). Insurance Design in the Presence of Integrated Medical Multiple Technologies. *American Economic Review 97*(2), 427–432.

Grossman, M. (1972). On the Concept of Health Capital and the Demand for Health. *Journal of Political Economy 80*(2), 223.

Harris, A. H. S., C. Chen, A. D. Rubinsky, K. J. Hoggatt, M. Neuman, and M. E. Vanneman (2016, April). Are Improvements in Measured Performance Driven by Better Treatment

or "Denominator Management"? *Journal of General Internal Medicine 31*(S1), 21–27.

Heckman, J. (2010, June). Building Bridges Between Structural and Program Evaluation Approaches to Evaluating Policy. *Journal of Economic Literature 48*(2), 356–398.

Hoerger, T., R. Harris, K. Hicks, K. Donahue, and S. Sorensen (2004). Screening for type 2 diabetes mellitus: a cost-effectiveness analysis. *Annals of Internal Medicine 140*(9), 689–699.

Huang, E. S., M. Shook, L. Jin, M. H. Chin, and D. O. Meltzer (2006, February). The impact of patient preferences on the cost-effectiveness of intensive glucose control in older patients with new-onset diabetes. *Diabetes Care 29*(2), 259–64.

Hyman, R. B., S. Baker, R. Ephraim, A. Moadel, and J. Philip (1994, August). Health Belief Model variables as predictors of screening mammography utilization. *Journal of Behavioral Medicine 17*(4), 391–406.

James, P., S. Oparil, B. Carter, W. Cushman, C. Dennison-Himmelfarb, J. Handler, D. Lackland, M. LeFevre, T. MacKenzie, O. Ogedegbe, S. Smith, L. Svetkey, S. Taler, R. Townsend, J. Wright, A. Narva, and E. Ortiz (2014, February). 2014 evidence-based guideline for the management of high blood pressure in adults: report from the panel members appointed to the Eighth Joint National Committee (JNC 8). *JAMA 311*(5), 507–20.

Kadiyala, S. and E. Strumpf (2016). How effective is population-based cancer screening? regression discontinuity estimates from the us guideline screening initiation ages. *Forum for Health Economics and Policy 19*(1), 87–139.

Kahn, R., P. Alperin, D. Eddy, K. Borch-Johnsen, J. Buse, J. Feigelman, E. Gregg, R. R. Holman, M. S. Kirkman, M. Stern, J. Tuomilehto, and N. J. Wareham (2010, April). Age at initiation and frequency of screening to detect type 2 diabetes: a cost-effectiveness analysis. *Lancet (London, England) 375*(9723), 1365–74.

Karve, A. M., F.-S. Ou, B. L. Lytle, and E. D. Peterson (2008, March). Potential unintended financial consequences of pay-for-performance on the quality of care for minority patients.

American Heart Journal 155(3), 571–576.

Kenkel, D. (1994). The demand for preventive medical care. *Applied Economics 26*(4), 313–325.

Kim, H. B., S. A. Lee, and W. Lim (2017). Knowing is Not Half the Battle: Impacts of the National Health Screening Program in Korea. *IZA diecussion paper series*.

Kim, H. B. and S.-m. Lee (2017). When public health intervention is not successful: Cost sharing, crowd-out, and selection in korea's national cancer screening program. *Journal of Health Economics 53*(Supplement C), 100 – 116.

Lange, F. (2011). The role of education in complex health decisions: Evidence from cancer screening. *Journal of Health Economics 30*(1), 43 – 54.

Lantz, P., D. Macklem, M. Hare, L. Richardson, L. Sever, and C. Orians (1997). Follow-up and treatment issues in the National Breast and Cervical Cancer Early Detection Program: results from a multiple-site case study - final report. *Baltimore Battelle, Centers for Public Health Research and Evaluation*.

Loeb, S., M. Bjurlin, J. Nicholson, T. L. Tammela, D. Penson, H. B. Carter, and R. Etzioni (2014). Overdiagnosis and overtreatment of prostate cancer. *European Urology 65*(6), 1046–1055.

Lostao, L., T. E. Joiner, J. W. Pettit, P. Chorot, and B. Sandín (2001). Health beliefs and illness attitudes as predictors of breast cancer screening attendance. *European Journal of Public Health 11*(3), 274–279.

Manning, W., J. P. Newhouse, N. Duan, E. B. Keeler, a. Leibowitz, and M. S. Marquis (1987, June). Health insurance and the demand for Medical Care: evidence from a randomized experiment. *The American economic review 77*(3), 251–77.

McDonald, M., R. Hertz, A. Unger, and M. Lustik (2009). Prevalence, awareness, and management of hypertension, dyslipidemia, and diabetes among United States adults aged 65 and older. *Journals of Gerontology - Series A Biological Sciences and Medical Sciences 64*(2), 256–263.

Meltzer, D. and E. Huang (2007). Effects of patient self-selection on cost-effectiveness of diabetes therapy by age. *iHEA 2007 6th World ...*, 2–3.

Musa, D., R. Schulz, R. Harris, M. Silverman, and S. B. Thomas (2009, Jul). Trust in the health care system and the use of preventive health services by older black and white adults. *American Journal of Public Health 99*(7), 1293–9.

Nathan, D. and W. Herman (2004). Screening for Diabetes: Can We Afford Not to Screen? *Annals of Internal Medicine 140*(9), 756–758+I6.

National Committee on Quality Assurance (2016). HEDIS Measures and Performance Measurement.

Olives, C., R. Myerson, A. H. Mokdad, C. J. L. Murray, and S. S. Lim (2013, January). Prevalence, awareness, treatment, and control of hypertension in United States counties, 2001-2009. *PloS one 8*(4), e60308.

Oster, E., I. Shoulson, and E. R. Dorsey (2013, April). Optimal Expectations and Limited Medical Testing: Evidence from Huntington Disease. *American Economic Review 103*(2), 804–830.

Patel, S., M. Winkel, M. Ali, K. M. V. Narayan, and N. Mehta (2015, June). Cardiovascular Mortality Associated With 5 Leading Risk Factors: National and State Preventable Fractions Estimated From Survey Data. *Annals of Internal Medicine 163*(4), 245–253.

Pauly, M. and F. Blavin (2008, December). Moral hazard in insurance, value-based cost sharing, and the benefits of blissful ignorance. *Journal of Health Economics 27*(6), 1407–17.

Rosenstock, I., V. Strecher, and M. Becker (1988, January). Social Learning Theory and the Health Belief Model. *Health Education and Behavior 15*(2), 175–183.

Sculpher, M. (2008, January). Subgroups and heterogeneity in cost-effectiveness analysis. *PharmacoEconomics 26*(9), 799–806.

Song, Z., S. Rose, D. G. Safran, B. E. Landon, M. P. Day, and M. E. Chernew (2014). Changes in health care spending and quality 4 years into global payment. *New England*

Journal of Medicine 371(18), 1704–1714. PMID: 25354104.

Song, Z., D. G. Safran, B. E. Landon, Y. He, R. P. Ellis, R. E. Mechanic, M. P. Day, and M. E. Chernew (2011). Health care spending and quality in year 1 of the alternative quality contract. *New England Journal of Medicine 365*(10), 909–918. PMID: 21751900.

Stone, N., J. Robinson, A. Lichtenstein, C. N. Bairey Mertz, C. Blum, R. Eckel, A. Goldberg, D. Gordon, D. Levy, D. Lloyd-Jones, P. McBride, J. S. Schwartz, S. Shero, S. Smith, K. Watson, and P. Wilson (2014, July). 2013 ACC/AHA Guideline on the Treatment of Blood Cholesterol to Reduce Atherosclerotic Cardiovascular Risk in Adults: A Report of the American College of Cardiology/American Heart Association Task Force on Practice Guidelines. *Journal of the American College of Cardiology 63*(25 Pt B), 2889–934.

The CDC Diabetes Cost-Effectiveness Study Group (1998). The Cost-effectiveness of Screening for Type 2 Diabetes. *JAMA 280*(20), 1757–1763.

The Center for the Evaluative Clinical Sciences, Dartmouth Medical School (1996-2017). The Dartmouth Atlas of Health Care: Data on Quality/Effective Care.

Vanness, D. and J. Mullahy (2012, March). Moving Beyond Mean-based Evaluation of Health Care. In *The Elgar Companion to Health Economics, Second Edition*, Chapters, Chapter 52. Edward Elgar.

Wang, C., A. Cheung, K. Bibbins-Domingo, L. Prosser, N. Cook, and L. Goldman (2011). Effectiveness and cost-effectiveness of blood pressure screening in adolescents in the United States. *Journal of pediatrics 158*, 257–64.

Wilson, N. (2011, August). Antiretroviral Therapy and Demand for HIV Testing: Evidence from Zambia. *SSRN Electronic Journal*.

Zweifler, R., L. McClure, V. Howard, M. Cushman, M. Hovater, M. Safford, G. Howard, and D. Goff (2011, January). Racial and geographic differences in prevalence, awareness, treatment and control of dyslipidemia: the reasons for geographic and racial differences in stroke (REGARDS) study. *Neuroepidemiology 37*(1), 39–44.