University of Virginia

A Retrospective Clinical Analysis of Prostate Cancer Screening and Outcomes in the UVA Health System: A Focus on Racial Disparities

By: Phillip Given

Economics DMP Thesis

Advisor: Lee Lockwood

4/15/20

ABSTRACT

Though racial disparities across all stages of prostate cancer management have been well established, uncertainty remains about the specific causes of these disparities and how to best address them. Significant controversy also exists regarding the efficacy of prostate cancer screening in general. This paper attempts to investigate these questions via logistic regression models using retrospective clinical data from the UVA Health System. I find that black men are significantly more likely to receive a biopsy at a given age and PSA result (O.R. = 1.90; $p < 0.001$), as well have being more likely to have prostate cancer at a given age/PSA result (O.R. = 2.39; $p < 0.001$), especially high-grade cancer. I also find that although black men have significantly higher odds of mortality overall, having prostate cancer does not appear to affect the odds of all-cause mortality in black men any more than white men, suggesting that large, general racial health disparities present a significant obstacle to achieving equal prostate cancer outcomes. Additionally, I explore regression discontinuity design as a possible quasi-experimental technique to measure screening efficacy without the need for a large-scale randomized controlled trial.

Acknowledgements: I would like to thank all of the terrific economics professors I have had the pleasure of knowing and learning from while at UVA. I especially thank Lee Lockwood for serving as my thesis advisor, as well as Amalia Miller for kindling my interest in the economics of healthcare and providing support in formulating this thesis. I also thank all of the physicians, scribes, nurses, and staff at Urology of Virginia for teaching me so much about urology, thus helping to inspire this thesis. I thank Tanya Wanchek, Ken Scully, Jason Lyman, Steve Patterson, and Ron Grider for help accessing UVA health system data. Most importantly, I would like to thank my parents for providing me with so much support and encouragement throughout this semester and my entire academic life; I could not have done any of this without you.

1. Introduction

Significant uncertainty exists about the efficacy of prostate cancer screening as well as the causes of racial disparities in prostate cancer outcomes. This paper attempts to investigate these questions using retrospective clinical data gathered from the University of Virginia Health System's Clinical Data Repository.

Since the widespread adoption of PSA screening to detect prostate cancer in the 1990s, there has been controversy about the risks and benefits of screening, who should be screened, and how much to weigh PSA results when making diagnostic and treatment decisions. I attempt to use a quasi-experimental regression discontinuity design to infer the causal effects of prostate cancer screening at a commonly used threshold for "Elevated PSA". However, contrary to previous research, I find no discontinuity in the probability a patient receives a biopsy at the threshold; instead, I see a continuous rise in the probability of biopsy as PSA increases, so I am unable to use this technique to make causal inferences on the effects of receiving a prostate biopsy on cancer detection and mortality.

Prostate cancer has long had a disproportionate impact on black men relative to white men. Black men are about 1.7 times more likely to develop prostate cancer than white men, and around 2.4 times more likely to die from it (Siegal et al. 2016). The causes for these disparities are thought to be complex and multifactorial, and include healthcare access and affordability, socioeconomic status, diet and comorbidities, and genetic/epigenetic factors. This paper investigates the effects of race on prostate cancer detection and mortality outcomes for patients in the UVA Health System using logistic regression models.

I find that black men are significantly more likely to receive a prostate biopsy and are also more likely to have a prostate cancer diagnosis, especially high-grade cancer, controlling for age and PSA result. Looking at all-cause mortality outcomes as of 2017, I find that black men have significantly higher mortality odds than white men, and that the effect of being black on mortality is even greater than the effect of having prostate cancer ($p < 0.01$). However, the results point to a prostate cancer diagnosis having a smaller effect on the mortality odds for black men than white men, suggesting that much of the disparity in prostate cancer mortality rate comes from general disparities between races in health and healthcare, and not factors specific to prostate cancer itself and the treatment thereof.

This paper contributes to the existing literature on the subject by exploring regression discontinuity as a plausible design for studying the efficacy of prostate cancer screening in recent retrospective clinical data, as opposed to relying solely on large-scale randomized controlled trials. Though I conclude that this technique is unlikely to be effective using modern clinical datasets of this size, it may be viable with a much larger group of patients. Additionally, I contribute to previous work on racial disparities in prostate cancer through the use of a multiple logistic regression model to try to isolate the effect of race on cancer detection and mortality, as opposed to simply reporting differences in population averages.

2. Background

2.1 Prostate Cancer and PSA

Prostate cancer continues to be a major cause of death for men; it is the second most common cancer for men, as well as the second most common cause of cancer death in men, behind only lung cancer (U.S. Cancer Statistics Working Group, 2019). Over the past few

decades, the rate of prostate cancer deaths has declined, partly due to improvements in screening and treatment technologies (National Cancer Institute, 2019). An elevated blood serum concentration of prostate-specific antigen (PSA), a glycoprotein enzyme that is produced both by cancerous and noncancerous prostate cells, is considered to be a sign of prostate cancer. The blood test for prostate-specific antigen is a very common screening technology for prostate cancer, with 38.8% of US men aged 55-69 receiving the test within the past year as of 2015 (National Cancer Institute, 2019). However, that statistic had declined almost 10% from 2008, when 48% of those men reported having been screened in the past year. This decline in PSA screening coincided with increasing concerns about the risks and benefits of widespread screening. The results of two large-scale, randomized trials in both the United States and Europe were published in 2009, and both showed little evidence of mortality benefits from broad PSA screening (Andriole et al. 2009; Schroder et al. 2009). Due to the financial costs and possible physical harm from overdiagnosis, many medical associations and panels, including the American Urological Association, relaxed their guidelines concerning PSA screening, restricting the age range and frequency for which PSA screening is recommended (Carter et al. 2013).

Commonly, a patient with a PSA concentration at or above 4.0 nanograms per milliliter of blood serum is considered to have an "elevated PSA". Since PSA typically rises as men age, a PSA of 4 would be more suspicious in a younger man than an older man, but 4.0 ng/mL is still widely considered as a cutoff for all men within the screening age range. Some men naturally have larger prostates and have higher PSA values as a result, so physicians also consider trends in PSA in addition to the level. An abrupt rise in PSA potentially indicates the presence of cancer (they also may consider measures such as PSA density, which is the PSA level divided by prostate volume, or %Free PSA, the percentage of PSA in blood that is not bound to a transport

protein). Despite these additional considerations, there still may exist a significant discontinuity in treatment decisions at a PSA level of 4.0 ng/mL. A patient with a PSA just above that threshold may be more likely to receive a prostate biopsy or MRI, be diagnosed with prostate cancer, and receive cancer treatment than those who fall just below the PSA threshold.

Like many other diseases, significant racial disparities exist in the incidence, screening, treatment, and mortality of prostate cancer between black and white men in the United States (Pietro et al. 2016). Many different mechanisms have been proposed to explain this disparity, including access to health care, socioeconomic status, cultural factors, diet and comorbidities, as well as genetic and epigenetic factors (Bhardwaj et al. 2017). It is likely that a combination of these factors, and not a single factor alone, causes the large disparity in prostate cancer outcomes.

2.2 Prior Literature

Regression discontinuity designs have become increasingly common in epidemiology and health research in recent years. Treatment decisions are often based on discrete cutoffs on a continuously measured variable, so researchers may compare patients who fall above and below a cutoff to make causal inferences (Bor et al., 2014). Assuming that treatment assignment is "as good as random" at values very close to the threshold, regression discontinuity can allow researchers to determine the causal effects of some medical interventions without the need for a randomized controlled trial. In economic literature, regression discontinuity has been used to estimate the value of receiving additional medical care as a result of being above or below a certain diagnostic threshold, such as those for low birthweights or a diabetes diagnosis (Almond et al. 2010; Alalouf et al. 2019). Using regression discontinuity designs in order to compare

differences in health outcomes with additional medical spending could allow healthcare providers to more efficiently allocate medical resources.

In spite of multiple, large-scale randomized trials, there continues to be uncertainty concerning the proper use of PSA screening, and physicians face conflicting guidelines about when to use this test. While the American Urological Association (AUA) used to recommend offering the test to all men over 50 with >10 year life expectancy, in 2013 the AUA updated its guidelines to recommend "shared decision making" for men aged 55-69, and does not recommend screening for men outside of that age range (Carter, et al. 2013). For comparison, the Canadian Task Force for Preventative Health Care does not recommend screening for men of any age (CTFPHC, 2014). However, amidst the recent plateau in prostate cancer deaths in the US (National Cancer Institute 2019), some have called for more screening. Screening advocates argue that the PSA test can be beneficial if implemented responsibly, and that new technologies, such as multi-parametric MRI, and better treatment strategies may help avoid overdiagnosis and reduce the possible harms from PSA screening (European Association of Urology, 2019).

In the United States, the Prostate, Lung, Colorectal, and Ovarian (PLCO) Cancer Screening Trial took place mostly in the 1990s and involved 76,693 men who were randomized into a screening group, which received annual PSA tests, and a control group. The control group was able to be screened outside of the study, resulting in a contaminated population and some concerns about the validity of the study. There was a higher incidence of prostate cancer found in the screening group, but there were no significant differences in mortality (Andriole, et al. 2009). The European Randomized Study of Screening for Prostate Cancer (ERSPC) took place concurrently and was roughly twice as large as the PLCO trial. The ERSPC trial found a small decrease in prostate cancer mortality in the screening group. The absolute risk difference was

0.71 deaths per 1000 men, meaning that 1410 men would need to be screened (and 48 extra cases of prostate cancer would need to be treated) in order to prevent one prostate cancer death (Schroder, et al. 2009).

Partially due to concerns of contamination in the control group of the PLCO trial, Shoag et al. published a regression discontinuity analysis of the PLCO data in 2015, which allowed them to evaluate the benefits of PSA screening without using data from the control group. Taking advantage of the fact that a PSA of 4.0 ng/mL was used as a cutoff for further work-up in the PLCO trial, they compared patients with maximum PSA values that fell just above and below the cutoff. They found that there was a positive effect on the probability of receiving a biopsy and detecting low-risk prostate cancer at the cutoff, but no effect on the detection of high-risk cancer or mortality. Surprisingly, they found a small, positive effect on prostate cancer-specific mortality at the cutoff, which is the opposite of what would be expected if screening were beneficial, though this was not statistically significant ($p = 0.27$).

Racial disparities have been noted across all stages of prostate cancer management, from screening and treatment choices to mortality outcomes (Pietro et al. 2016). A 2014 study by Powell et al. using SEER data found that racial disparities in age-adjusted 5-year survival rates for all stages of prostate cancer were significantly reduced in the PSA screening era (1995-2005) compared to the pre-PSA era (1973-1994), suggesting that aggressive PSA screening can improve outcomes in black men relative to white men. Unfortunately, there is evidence that black men have lower PSA screening rates despite having higher incidence and mortality (Trantham et al. 2013). This discrepancy could be due to differences in health care access and affordability, as well as skepticism within the black community towards doctors and hospitals

due to a history of unequal treatment and segregation, as well as a cultural memory of harmful and prejudicial events such as the Tuskegee Syphilis Study (Moses and Brantley, 2020).

Heterogeneity in forms of treatment exists between races as well. From SEER data, black men are less likely to receive radical prostatectomy, brachytherapy, cryotherapy, or combination therapy relative to white men, with more black men receiving active surveillance/watchful waiting (Moses et al. 2016). Although a more equal provision of healthcare can result in a narrowing of differences in outcomes, black men may still experience higher incidence and mortality due to genetic factors. A number of gene polymorphisms and mutations has both been associated with increased prostate cancer incidence and aggressiveness, and are found more frequently in black men. Other biological risk factors have also been reported to be more prevalent in black men, including differences in epigenetics, microRNA expression, and hormone-receptor, growth factor receptor, and inflammatory signaling pathway activation. Much of the relative importance of each of these factors is still unknown (Bhardwaj et al. 2017).

3. Data

All data was gathered from the University of Virginia's Clinical Data Repository (CDR). The CDR is a data warehouse that integrates and stores data gathered from multiple different sources across the UVA Health System (Einbender et al. 2001). Original data sources pertinent to this paper include electronic health records for demographic information, SunQuest for laboratory results, UVA's Cancer Tumor Registry, and the Virginia Department of Health for death information. The CDR contains data from 1.5 million patients in total, beginning in 1993. However, pathology results were not included until 2005; therefore, the cohort used in this paper

only contains patients with PSA results from 2005 and later, in order to avoid patients with unrecorded prostate biopsies. The CDR stopped updating new data in 2017.

In order to protect patient anonymity, specific dates are not included in the CDR. Some information, such as laboratory results, includes the year of the test. However, no date is recorded with pathology results, and therefore I was not able to link biopsies to the specific PSA result that triggered the biopsy.

To build the cohort, I pull male patients aged 40 and above with at least one PSA result on file between the years 2005-2017, which leaves 43,394 patients. Since, the CDR only contains data from the UVA Health System, some results from patients who also receive care from other health care providers may not have been recorded in the CDR. In order to avoid those cases in which the patient is missing important PSA results, I excluded patients with no PSA results greater than or equal to 1.0 ng/mL, leaving 21,985 patients and a total of 66,573 PSA tests $\geq$ 1.0 ng/mL. This exclusion reduced the number of patients in the sample with prostate cancer from 2,501 to 1,279. Since it would be very unlikely to perform a biopsy and find cancer in a patient without them ever having a PSA above 1, The large number of patients with prostate cancer without a PSA $\geq$ 1 is evidence that there is either incomplete data for these patients and/or there are patients with a history of prostate cancer prior to 2005 who have already been treated, and therefore have later PSA results that are very low. For either of these cases, I would want to exclude these patients in order to ensure the sample contains mostly patients in which the PSA result that triggered the initial biopsy is recorded.

To perform the first-stage regression discontinuity analysis, I estimate the change in probability of having received a prostate biopsy at the elevated PSA cutoff of 4.0 ng/mL. For this analysis I use each patient's maximum PSA result, with the assumption that that result triggered

the biopsy (This is the same assumption made in Shoag et al. 2015). Figure 1 presents a histogram of maximum PSA results from 1.0 to 20.0 ng/mL. Table 1 shows descriptive statistics for the sample. Average age is 65.06 years old, 77.9% are white, 12.1% are black, 11.4% have had a biopsy, and 5.8% have had a prostate cancer diagnosis. As expected, the histogram of each patient's maximum PSA result is heavily right skewed, with a mean of 19.88 ng/mL and a median of 2.37 ng/mL, and there are several outlier PSA values well above 1000 ng/mL. Results are consistent with or without inclusion of these outliers.

4. Empirical Strategy

I use a regression discontinuity design to estimate the change in probability of receiving a prostate biopsy at the PSA threshold of 4.0 ng/mL. For this analysis, I use local linear regression and the Calonico et al. 2018 MSE-optimal bandwidth selector. An implicit assumption in regression discontinuity frameworks is that data cannot be manipulated by patients or providers in order to be on a preferred side of the threshold. In this setting, manipulation is unlikely by either party due to its nature as a blood test and the fact that treatment is not restricted above or below the threshold. It is theoretically possible that patients could attempt to increase their PSA value via prostate stimulation from sexual activity or long-distance bicycle riding, but they would have no clear incentive to do so. Clumping of values on either side of the threshold is also not seen in Figure 1, as the density is smooth across the threshold. The regression equation can be expressed as follows:

$$\boldsymbol{T_i = \gamma + \delta D_i + f(X_i - 4) + v_i}$$

$\boldsymbol{T}$ = Probability of undergoing a biopsy
$\boldsymbol{D}$ = 1 if PSA ≥ 4.0; 0 if PSA < 4.0
$\boldsymbol{X}$ = Max PSA value

$\boldsymbol{v}$ = error term

Here, the coefficient of interest is $\boldsymbol{\delta}$, which is the estimate of the effect of being above the threshold on the probability of undergoing a biopsy. A statistically significant positive effect would then allow for further analysis of the effect of receiving a biopsy on cancer detection and mortality through a 2nd stage regression. In this case, the PSA threshold would serve as an instrumental variable for receiving a biopsy.

For my analysis of racial disparities in cancer detection and mortality outcomes I utilize logistic (or logit) regression, as all the dependent variables used are binary. I report results from logistic regression as opposed to a probit model due to ease of interpretation of coefficients and to stay consistent with other epidemiological studies. My results are consistent across both models. The basic logit model specification could be expressed as:

$$\boldsymbol{logit}(\boldsymbol{p_i}) = \boldsymbol{ln}\left(\frac{\boldsymbol{p_i}}{\boldsymbol{1-p_i}}\right) = \boldsymbol{\beta_0} + \boldsymbol{\beta_1 R_i} + \boldsymbol{B_X X_i} + \boldsymbol{\epsilon_i}$$

$\boldsymbol{p}$ = P(Y=1), where Y is the binary outcome variable (e.g. biopsy)
R = 1 if the patient is black
X = a vector of control variables

This model is extended by including interaction terms to determine differences in the effect of one independent variable based on the level of a second independent variable, such as the differences in the effect of prostate cancer on mortality across different races.

5. Results

5.1 Regression Discontinuity

Using each patient's maximum PSA result between the years 2005 and 2017, I performed a regression discontinuity analysis to determine if there is a jump in the probability of receiving a biopsy between those same years at the elevated PSA cutoff of 4 ng/mL. Contrary to the Shoag et al. 2015 findings using PLCO data, I find no statistically significant discontinuity at the cutoff (point estimate = -0.0123; p = 0.702; 95% CI, -0.0755 – 0.0508), and this result is consistent across a variety of bandwidths and specifications. Figure 2 displays the fitted relationship between PSA result and probability of biopsy. For PSA results below 4 ng/mL, the probabilities of receiving a biopsy are very similar to those in the PLCO data, at around 5% for a PSA = 2 ng/mL and around 10% for a PSA approaching 4 ng/mL. However, on the other side of the threshold, the probability immediately jumped to around 35% in the PLCO data, while my results show a gradual rise in probability with no discontinuity.

The reason for this major difference is likely that in the PLCO trial a PSA of 4.0 ng/mL was officially set as the threshold for further workup (i.e. either a biopsy or a follow-up PSA test), while in my sample no such official protocol exists. While there still are some loose guidelines, patients and doctors have more flexibility in their screening decisions than in the PLCO trial. It is also likely that the results of the PLCO and other similar trials, which showed few benefits to screening, resulted in physicians being less likely to recommend biopsies to patients with PSA results just above 4 ng/mL, and paying less attention to the 4.0 ng/mL mark as a discrete cutoff on which to base their screening decisions. It is also a possibility that a small, positive discontinuity does exist in the current population, but my sample size (N = 21,985) was not large enough to detect it. A much larger sample would most likely be needed to detect a statistically significant discontinuity in the current population, if it exists.

Without a discontinuity in the 1st stage regression (probability of biopsy), I cannot determine a causal estimate of the effect of a biopsy on cancer detection or mortality. Unsurprisingly given the lack of discontinuity in the treatment variable (biopsy), there is also no discontinuity in the probability of a prostate cancer diagnosis or in mortality at the PSA cutoff, as there is no reason such discontinuities would exist except as a result of increased probability of biopsy at the cutoff. Compared to the results from Shoag et al. 2015, the lack of a positive jump in both biopsies and low/intermediate-grade tumors at the PSA cutoff (**Figure 2)** suggests that the reduction in the number of biopsies in the PSA range just above 4 ng/mL in the current population may have led to fewer low-grade cancer diagnoses for patients with PSA results in that range. Given previous research (Andriole et al. 2009; Schroder et al. 2009; Shoag et al. 2015) showing little to no gains in survival as a result of PSA screening due to the low mortality risk from many low-risk prostate cancer cases, such a reduction in low-risk diagnoses may have had a beneficial effect in terms of cost-effectiveness and even mortality itself, though further, experimental research would of course be necessary to confirm this effect.

5.2 Racial Disparities

Focusing on racial disparities between black and white men, I estimate the effect of race on different outcome variables using logistic regression models. Note that by "effect" I do not mean the strict causal effect from race, since I was not able to control for all other variables correlated with race that may also have an effect on outcomes, such as socioeconomic status or family history of prostate cancer. Rather, these regression coefficients are simply estimates of the ratios of population odds between races, controlling for certain variables such as age or max PSA result.

First, I examine disparities in the probability of undergoing a biopsy or receiving a prostate cancer diagnosis, controlling for max PSA result, age, and the interaction term between the two (**Table 2**). The effect of a PSA difference on the likelihood of biopsy may decrease as patient age increases, since it is normal for patients to have a higher PSA as they age and physicians are more hesitant to biopsy older patients (e.g. >70-75 years old) due to a lower chance of prostate cancer mortality, so a difference in PSA at a higher age may have a smaller effect on probability of receiving a biopsy than at a lower age. This effect is seen by an odds ratio of <1 for the interaction term in table 2. The odds ratio >1 for age on probability of biopsy may seem contradictory if physicians are less likely to biopsy older patients at a given PSA, but that can be explained by the fact that the outcome variable is the probability of ever receiving a biopsy, which clearly increases with age, not the probability of receiving a biopsy just after that specific PSA result.

In these regressions, we see that black patients are much more likely to have received a biopsy controlling for age and max PSA result (O.R. = 1.90; 95% CI, 1.70 – 2.13), suggesting that physicians and patients may be taking race into account when making treatment decisions and recommending biopsies more aggressively to black patients due to higher prostate cancer incidence and mortality rates. Indeed, black patients are far more likely to have had prostate cancer (O.R. = 2.39; 95% CI, 2.07 – 2.75) again controlling for age and PSA, suggesting that more biopsies are well merited. It is possible that the higher incidence of prostate cancer is caused simply by more biopsies being performed on black men, and therefore more low-grade tumors would be detected. However, the fact that the odds ratio for high grade tumors (O.R. = 2.45) is even greater than that of low/intermediate grade tumors (O.R. = 2.15) suggests that that is not the case.

Next, I examine the effects of race and prostate cancer status on all-cause mortality at the exit date of 2017, controlling for year of birth (**Table 3**). I find that black patients in the sample, controlling for whether or not they have prostate cancer, have much higher odds of mortality (O.R. = 1.89; 95% CI, 1.67 – 2.14), and that patients with prostate cancer have higher odds of mortality (O.R. = 1.43; 95% CI, 1.23 – 1.67) controlling for race. The higher risk of mortality appears to come entirely from patients with high grade tumors (O.R. = 1.75; 95% CI, 1.42 – 2.16), while low/intermediate tumors have no statistically significant effect on mortality. Notably, the effect of being black on mortality is greater than the effect of having prostate cancer ($p < 0.01$), and the difference in the effects of having high-grade cancer and being black is not statistically significant. Figure 3 displays the Nelson-Aalen cumulative hazard curves across race and cancer status for ages 60-80.

I add interaction terms between race and prostate cancer status in order to examine if the effects of having prostate cancer on mortality varies between races. Again controlling for year of birth, prostate cancer appears to have a smaller effect on mortality for black men (O.R. = 0.72, 95% CI, 0.50 – 1.04, $p = 0.08$), though this difference is only statistically significant at the 8% level. Among only black men, there is no significant effect on all-cause mortality from having prostate cancer (O.R. = 1.13; 95% CI, 0.82 – 1.56; $p = 0.466$), although one potential concern with this estimate is lack of statistical power due to a much smaller sample size of black men (N = 2,659) than that of all men. Among patients of any race with prostate cancer, being black still increases risk of all-cause mortality (O.R. = 1.43, 95% CI, 1.01 – 2.05; $p = 0.44$).

In summary, the racial disparity in mortality among prostate cancer patients in this population appears to come from the large, systemic health disparity between black and white men, rather than from idiosyncratic differences in how prostate cancer effects mortality for each

race. This evidence suggests that the overall racial mortality gap presents a significant obstacle to closing the gap among prostate cancer patients specifically. Even if prostate cancer has a smaller effect on mortality for black men than white men, it may be outweighed by the underlying mortality risk from being black. In fact, it is possible that a prostate cancer diagnosis may act as a sort of "equalizing" factor between a given black and white patient's expected mortality, since although it effects each person's expected survival negatively, it may have a larger effect on the white patient, because he would otherwise outlive the black patient even longer on average.

6. Conclusion

Prostate cancer continues to disproportionately affect black men, both in terms of incidence and prostate cancer mortality. This paper attempts to explore the effect of race on prostate cancer biopsies, incidence, and outcomes using data from the UVA Health System. Instead of simply reporting differences in population averages like most previous epidemiological work on this subject, I attempt to isolate the effect of race from other correlated variables (i.e. PSA result and prostate cancer status) in order to more accurately measure its effect on screening and mortality outcomes. I find that black men are significantly more likely to have received a biopsy at a given PSA result. They are also significant more likely to have a prostate cancer diagnosis controlling for patients' maximum PSA value, especially high-grade prostate cancer, suggesting that an aggressive screening approach for black men relative to white men is merited.

Interestingly, using data on all-cause mortality as of 2017 I also find that a prostate cancer diagnosis has a smaller effect on the odds of mortality for black men in the sample

relative to white men, though this difference is only statistically significant at the level $\alpha = .08$. However, black men with prostate cancer still have higher odds of mortality than white men with prostate cancer, due to the large mortality gap between black and white men in general. This evidence suggests that even with equal prostate cancer-specific treatment, black men will still face worse prostate cancer mortality outcomes due to a higher incidence and worse overall mortality. In order to close the mortality gap among prostate cancer patients, significant improvements in overall racial health disparities will have to be made. From a more optimistic perspective, the evidence suggests that with proper screening and treatment, a prostate cancer diagnosis does not have to affect the odds of mortality for a black patient any more than it does for a white patient.

Several limitations exist in this paper, especially due to the nature of the data used. I rely on all-cause mortality, so I am not able to differentiate those patients who die *from* prostate cancer from those who die with a prostate cancer diagnosis, whether or not prostate cancer was responsible for the death. Another concern is the potential for incomplete data, since data may be missing for some patients who received medical services from other healthcare providers. Finally, since the data comes from only one healthcare provider that mainly operates in one locality (Charlottesville, VA and the surrounding area), the sample used in this paper may not be representative of the broader US population, due to potential differences in patient and provider characteristics.

Future research is needed to evaluate the efficacy of current prostate cancer screening protocols and the causal effects of race on prostate cancer outcomes. Although fruitless in this paper, it may be possible to identify and exploit a discontinuity at a PSA cutoff point in order to make causal inferences concerning current screening effectiveness by using a much larger and

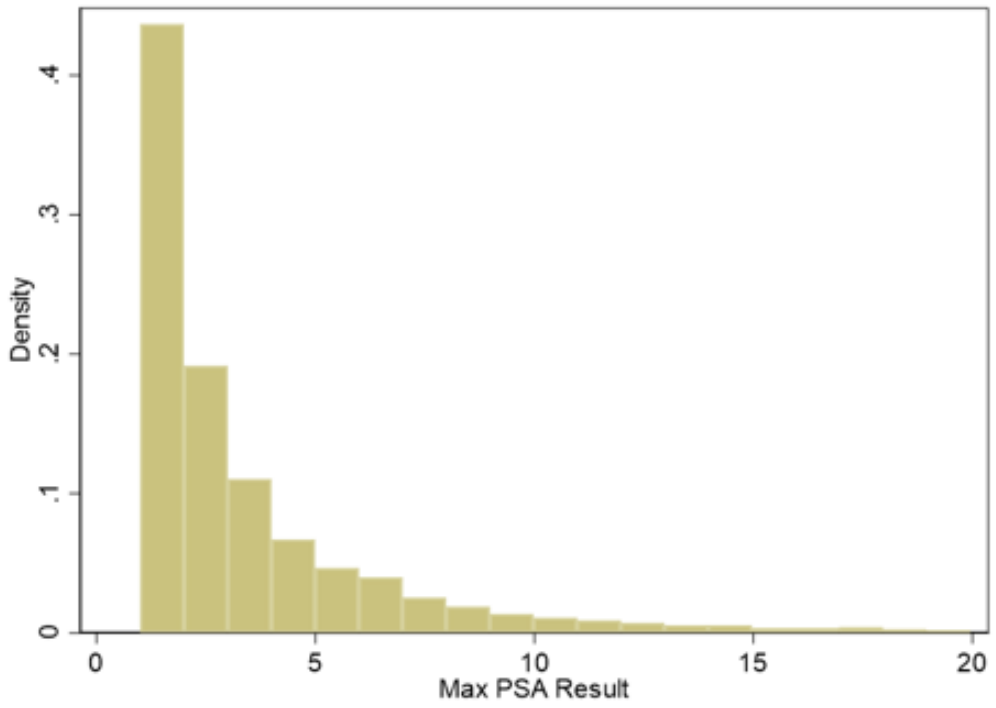
more complete dataset. To further investigate racial disparities in prostate cancer, future research should utilize multiple regression models with numerous controls, in addition to experimental or quasi-experimental analyses, in order to identify the root causes of the disparities and hopefully make further progress towards equal outcomes.

Table 1: Descriptive Statistics of CDR data

	Mean
Age	65.061
White	0.779
Black	0.121
Other Race	0.100
Deceased	0.137
Max PSA	19.880
Biopsy	0.114
Tumor	0.058
High Grade	0.027
Low/Int Grade	0.028

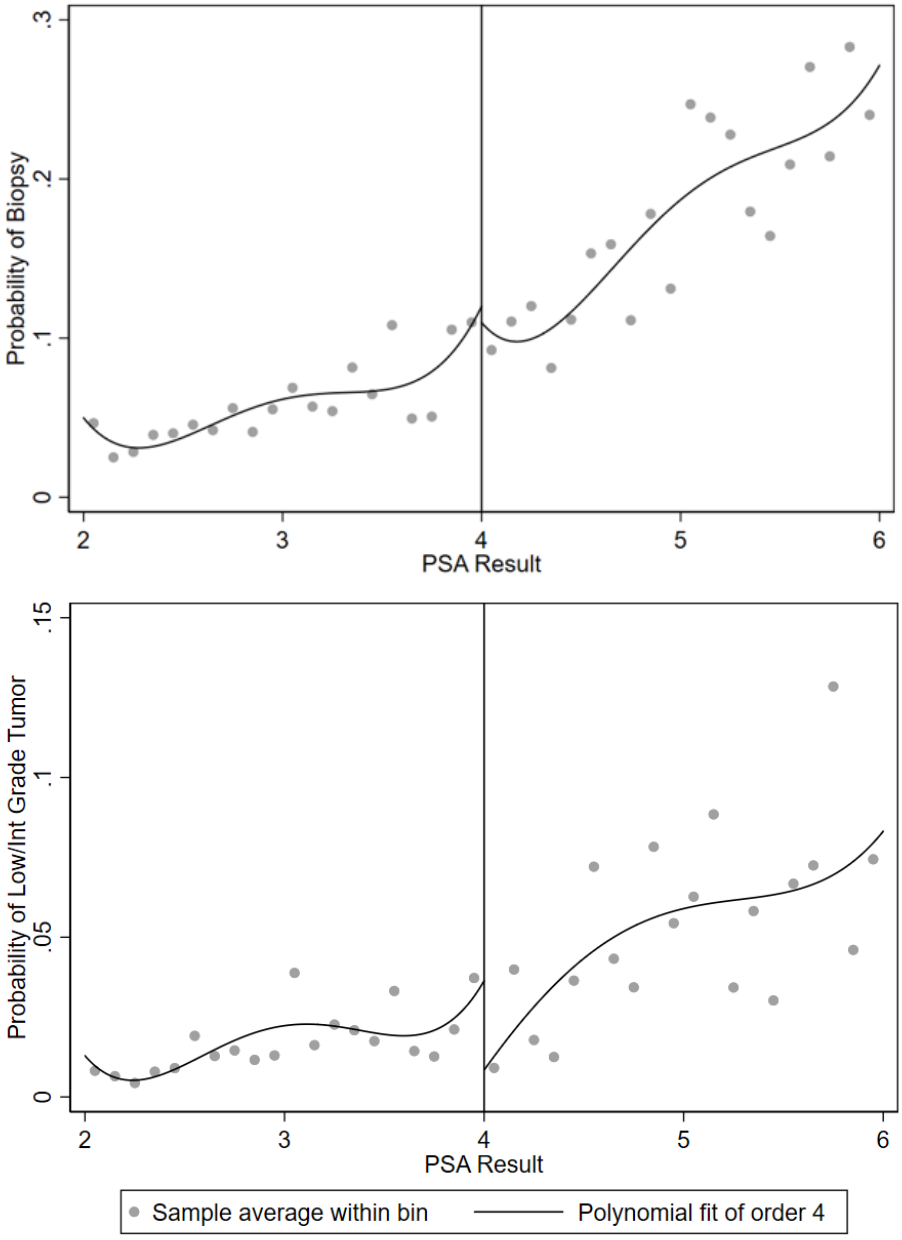
Note: Author's calculations from UVA CDR data. This table contains descriptive statistics for patients with a PSA test who meet all sample criteria described in the text. "Age" indicates age at time of maximum PSA test. "Deceased" = 1 if patient is dead as of exit date of 2017. Tumors were categorized by grade as reported in CDR data from the Cancer Tumor Registry, with grades 1-2 being "Low/Int Grade" and grades 3-4 being "High Grade".

Figure 1: Histogram of Maximum PSA results



Note: Horizontal axis displays patients' maximum PSA result in ng/mL. Vertical axis displays proportion of total patient sample with a maximum PSA result with each bin.

Figure 2: Regression Discontinuity Analysis



Note: Horizontal axis displays patients' maximum PSA result in ng/mL. Vertical axis displays within PSA values probability of the outcome variable. Vertical line at 4 is the cutoff for an "elevated PSA".

Table 2: Effects on Probability of Biopsy and Prostate Cancer, as Odds Ratios (O.R.)

	Biopsy		Tumor		High Grade		Low/Intermediate grade	
	O.R.	p-value	O.R.	p-value	O.R.	p-value	O.R.	p-value
PSA result	1.0053	<0.001	1.0036	<0.001	1.0022	<0.001	1.0000	0.967
age	1.0204	<0.001	1.0290	<0.001	1.0232	<0.001	1.0264	<0.001
result · age	0.9999	<0.001	0.9999	<0.001	0.9999	0.003	1.0000	0.935
black	1.9046	<0.001	2.3880	<0.001	2.4544	<0.001	2.1545	<0.001

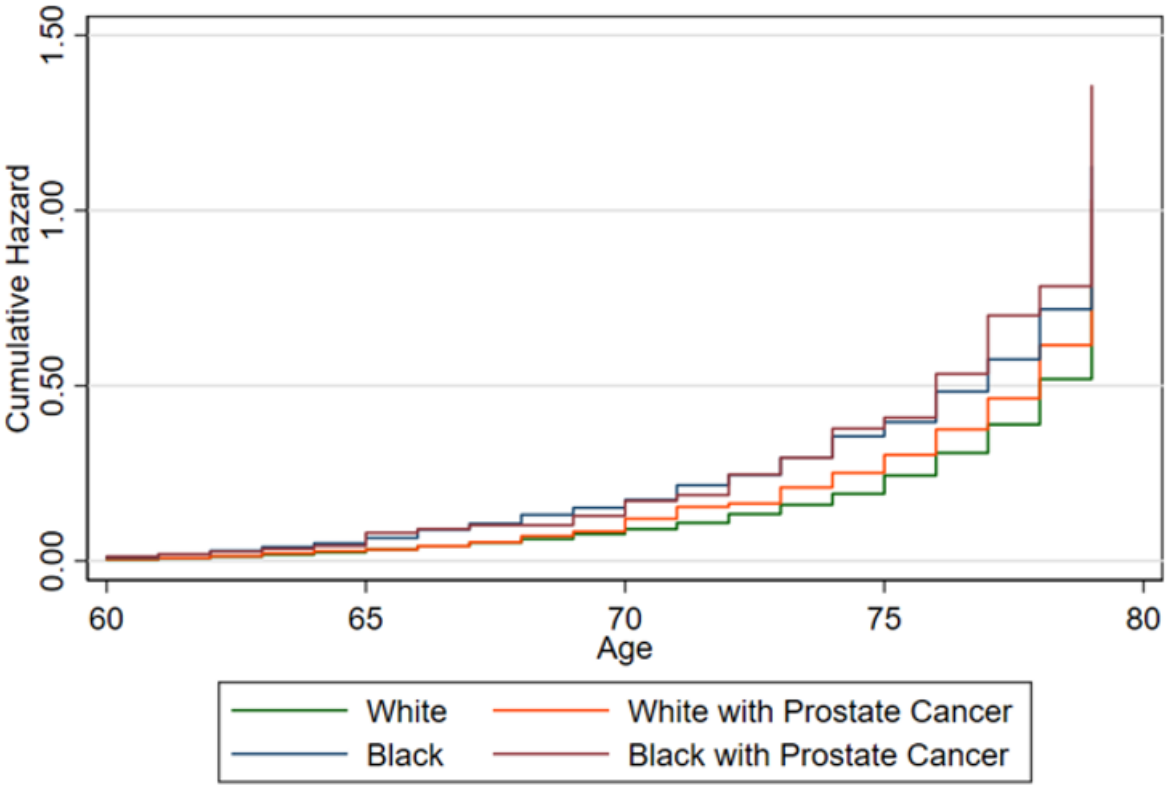
Note: black = 1 if patient is black, otherwise = 0. All p-values reported are two-sided. "Age" indicates patients age at time of their maximum PSA result.

Table 3: Effects on All-Cause Mortality as of 2017

	Specification							
	(1)		(2)		(3)		(4)	
	O.R.	p-value	O.R.	p-value	O.R.	p-value	O.R.	p-value
year of birth	0.8986	<0.001	0.8982	<0.001	0.8986	<0.001	0.8981	<0.001
tumor	1.4314	<0.001			1.5402	<0.001		
high grade			1.7505	<0.001			1.9560	<0.001
low/int. grade			0.8768	0.280			0.9395	0.645
black	1.8923	<0.001	1.9011	<0.001	1.9725	<0.001	1.9901	<0.001
black · tumor					0.7179	0.080		
black · high grade							0.6406	0.076
black · low/int. grade							0.7077	0.257

Note: black = 1 if patient is black, otherwise = 0. All p-values reported are two-sided.

Figure 3: Nelson-Aalen Cumulative Hazard Estimates



Note: Horizontal axis displays patient age. Vertical axis displays cumulative hazard. Cumulative hazard at age = x is calculated by integrating the hazard function over all ages ≤ x. The hazard function at age = x is equal to the probability a patient of age = x dies at that age. "White" and "Black" indicate white and black men respectively who do not develop prostate cancer, and "White with Prostate Cancer" and "Black with Prostate Cancer" include only men who develop prostate cancer. Nelson-Aalen curves were chosen to display survival analysis as opposed to Kaplan-Meier curves due to the extent of right-censored data in the sample.

References

Alalouf, M., S. Miller, and L. Wherry (2019, Oct) What Difference Does a Diagnosis Make? Evidence from Marginal Patients. *NBER Working Paper Series* No. 26363

Almond, D., J. J. Doyle, A. E. Kowalski, and H. Williams (2010, May). Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns. *The Quarterly Journal of Economics* 125(2), 591–634

Andriole GL, Crawford ED, Grubb RL, 3rd, et al. (2009) Mortality results from a randomized prostate-cancer screening trial [published erratum in *N Engl J Med*. 2009;360:1797] N Engl J Med. 2009;360:1310–1319.

Bhardwaj, A., Srivastava, S. K., Khan, M. A., Prajapati, V. K., Singh, S., Carter, J. E., & Singh, A. P. (2017). Racial disparities in prostate cancer: a molecular perspective. *Frontiers in bioscience (Landmark edition)*, *22*, 772–782.

Bor J, Moscoe E, Mutevedzi P, Newell M-L, Bärnighausen T. (2014). Regression discontinuity designs in epidemiology: causal inference without randomized trials. *Epidemiology* 25:729-37.

Calonico S., M. D. Cattaneo, and M. H. Farrell. (2018). Optimal Bandwidth Choice for Robust Bias Corrected Inference in Regression Discontinuity Designs.

"Canadian Task Force Recommends against Screening for Prostate Cancer." *Canadian Task Force on Preventive Health Care*, CTFPHC, 27 Oct. 2014.

"Cancer of the Prostate - Cancer Stat Facts." *SEER*, National Cancer Institute, 2019.

Carter, H. Ballentine, et al. (2013) *Prostate Cancer: Early Detection Guideline*. American Urological Association.

Einbinder J.S., K. Scully, R. Pates, J. Schubart, R. Reynolds (2001). Case study: a data warehouse for an academic medical center. *Journal of Healthcare Information Management*. 15(2):165-75.

"EPAD19: European PSA Screening Programme Is on Its Way - Part 2." *Uroweb*, European Association of Urology, 4 Feb. 2019, uroweb.org/epad19-european-psa-screening-programme-is-on-its-way-part-2/.

Moses, K. A., and J. B. Thrasher (2020). "Addressing Prostate Cancer's Racial Disparity Starts with You." *Urology Times*, MJH Life Sciences, 7 Feb. 2020.

Moses, K. A., Orom, H., Brasel, A., Gaddy, J., & Underwood, W., 3rd (2016). Racial/ethnic differences in the relative risk of receipt of specific treatment among men with prostate cancer. *Urologic oncology*, *34*(9), 415.e7–415.e12.

Pietro, G. D., Chornokur, G., Kumar, N. B., Davis, C., & Park, J. Y. (2016). Racial Differences in the Diagnosis and Treatment of Prostate Cancer. *International neurourology journal*, *20*(Suppl 2), S112–S119.

Powell, I. J., Vigneau, F. D., Bock, C. H., Ruterbusch, J., & Heilbrun, L. K. (2014). Reducing prostate cancer racial disparity: evidence for aggressive early prostate cancer PSA testing of African American men. *Cancer Epidemiology, Biomarkers & Prevention*, *23*(8), 1505–1511.

"Prostate Cancer Screening." *Prostate Cancer Screening*, National Cancer Institute, 2019.

Schroder FH, Hugosson J, Roobol MJ, et al. Screening and prostate-cancer mortality in a randomized European study. *N Engl J Med.* 2009;360:1320–1328.

Siegel, R.L., Miller, K.D. and Jemal, A. (2016), Cancer statistics, 2016. *CA: A Cancer Journal for Clinicians*, 66: 7-30.

Shoag J, Halpern J, Eisner B, et al. Efficacy of Prostate-Specific Antigen Screening: Use of Regression Discontinuity in the PLCO Cancer Screening Trial. *JAMA Oncol.* 2015;1(7):984–986.

Trantham, L. C., Nielsen, M. E., Mobley, L. R., Wheeler, S. B., Carpenter, W. R., & Biddle, A. K. (2013). Use of prostate-specific antigen testing as a disease surveillance tool following radical prostatectomy. *Cancer*, *119*(19), 3523–3530.

U.S. Cancer Statistics Working Group. U.S. Cancer Statistics Data Visualizations Tool, based on November 2018 submission data (1999-2016): U.S. Department of Health and Human Services, Centers for Disease Control and Prevention and National Cancer Institute; www.cdc.gov/cancer/dataviz, June 2019.