SUPPLEMENTARY MATERIAL

Data source

Data for this study are captured by the 18 registries comprising the Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI).[1] The SEER data contain diagnostic information on all tumors diagnosed within the catchment areas of Connecticut, New Mexico, Utah, Hawaii, Iowa, New Jersey, Kentucky, Louisiana, Georgia, California, Atlanta, Detroit, San Francisco-Oakland, San Jose-Monterey, Los Angeles, Seattle-Puget Sound, and among Arizona Native Americans, and Alaska natives. The 18 SEER registries, which cover approximately 28% of the US population and link to the National Center for Health Statistics.[2] This study includes all males over age 40 (at time of diagnosis) with a PCa diagnosis from 2004–2012 (N = 514,878).

County-level health care data were obtained from the Department of Health and Human Services Area Health Resource File (AHRF), a database that contains county-specific health care and economic measures (http://ahrf.hrsa.gov). These data include the number of physicians by subspecialty within a county obtained from the 2005 American Medical Association Physician Master files and the number of hospitals from the American Health Association Hospital Facilities Database. Males were linked to county-level data in AHRF using their county of residence at time of diagnosis.

Measures

Individual-level variables were derived using the SEER data. Time was measured as months from PCa diagnosis to PCa death and observations were treated as censored at time of death from all other causes or end of the follow-up period. Race/ethnicity was grouped into the following categories; African-American, non-Hispanic white, Hispanic, other race, or unknown race/ethnicity. Staging was coded using the SEER Derived American Joint Committee on Cancer 6th edition clinical stage from 2004 to the present.[1] Stage is based on information collected under the Collaborative Stage Data Collection

System (CS) and coded using the CS algorithm. AJCC Stage I diagnoses were relatively rare, accounting for 0.1% of all diagnoses, and were combined with Stage II diagnoses. Individuals with unknown stage at diagnosis were excluded from the analysis, as the factors leading to an unknown stage were heterogeneous and would yield no prognostic factors to affect clinical decision making (Figure 1). Tumor grade, established based on the SEER histologic grading system, was measured using the following categories; <= 6, 7 (combination of 3,4), 7 (combination of 4,3), 8, 9 & 10, and unknown.[3] Prostate screening antigen at time of diagnosis (PSA) values were grouped into categories: <4ng/mL, 4-10ng/mL, 10-20ng/mL, and >20ng/mL.

Selected county-level measures from the AHRF database included number of physicians and surgeons, radiation oncologists, urologists, and short-term general hospitals per 100,000 population. We expected a non-linear relationship between physician density and PCa mortality; therefore, we compared counties in the 25th and 75th percentiles to the middle 50% of the distribution. Based on findings from other studies [4, 5], we compared counties with one or more urologists to those with none. The AHRF also includes a code for Metropolitan/Micropolitan Statistical Areas, with the following categories; 1) rural, 2) Metropolitan Statistical Areas having at least one urbanized area of 50,000 or more population, and 3) Micropolitan Statistical Areas having at least one urban cluster with a population of 10,000–50,000. We expected a non-linear relationship between physician density and PCa mortality; therefore, we compared counties in the 25th and 75th percentiles to the middle 50% of the distribution.

The Agency for Heathcare and Research Quality (AHRQ) Prevention Quality Indicator (PQI) was used to measure the quality of care for "ambulatory care sensitive conditions." AHRQuality Indicators™. Content last reviewed July 2018. Agency for Healthcare Research and Quality, Rockville, MD. https://www.ahrq.gov/cpi/about/otherwebsites/qualityindicators.ahrq.gov/qualityindicators.html This set of measures can be used with hospital inpatient discharge data to identify quality of care for "ambulatory care sensitive conditions." The PQIs are population based and adjusted for covariates. Even though these indicators are based on hospital inpatient data, they provide insight into the

community health care system or services outside the hospital setting. For example, patients with diabetes may be hospitalized for diabetic complications if their conditions are not adequately monitored or if they do not receive the patient education needed for appropriate self-management.

County-level demographics from the 2000 U.S. Census were also included in the models. Median–family-income (MFI) and the percent of families below the poverty line were also drawn from the Census data. As this population is largely over age 65, we considered the proportion of the Medicare population in each county that is also eligible for Medicaid enrollment ("dual enrollees").[6] The Center for Disease Control (CDC) Social Vulnerability Index (SVI) was included in the models. The SVI uses 15 variables from the 2000 US Census data to measure four domains of social vulnerability; 1) Socioeconomic status (income, poverty, employment, and education), 2) Household composition and disability (age, single parents, and disability), 3) Minority status and language profile (race, ethnicity, and English language), and 4) Housing and transportation profile (housing structure, crowding, and vehicle access). Counties are then given an overall vulnerability rank.

Statistical Methods

Classification and Regression Trees

Traditional regression approaches specify interactions to test for differences in the effect of individual, family history, and neighborhood characteristics on prostate cancer mortality. However, there are some limitations to using interactions to bin individuals into groups with similar prognostic outcomes. All interactions need to be specified a priori and the number of interactions and interactions involving multiple variables are difficult to interpret. A classification and regression tree (CART) is an alternative method that allows us to explore the structure of the data with the goal of predicting survival outcomes based on individual and county level characteristics that may affect prostate cancer mortality.

A regression tree is a hierarchal structure that has a top node (or root) and observations are passed down the tree. Each decision point, which is selected by the algorithm to explain the most deviance, is labeled a node (sometimes called daughter nodes) until it reaches a terminal node (or leaf). CART uses a binary splitting process to identify the best model for classifying individuals into distinct groups. The central aim of the regression tree approach is to form meaningful classes that are determined by the data (not a priori assumptions). The results from CART likely do not represent additive functions that consist only of main effects, but complex interactions between variables in the data. Essentially, we are asking how a compilation of variables come together to define distinct sub-classes of individuals.

Random Forest

Random forests (RF) are an extension of CART. In this method, *n* trees are grown using a bootstrapped sample from the learning sample.[7, 8] The number of trees grown is specified by the user, with a default of 1000 in the R statistical package *randomForestSRC* with a logrank splitting rule. We chose to have the algorithm fit 200 trees and constrained the model to have a terminal node size of 50. Unlike CART, there is no trimming or stopping criterion, the trees are fully grown (the user can modify this criterion, however the standard practice is to fully grow the tree). Additionally, a subset of variables are randomly selected for inclusion at each node. This method of random subspace selection is done to avoid correlation between trees in the forest and decreases the error. It also allows for the selection of the most relevant variables when there are multicollinearity issues and therefore reduces the variables of interest to those with the most explanatory value.[9] All models had error rates that ranged between 17% and 29%, with the highest stage of disease models having the highest error rates.

One benefit of the RF method is the ability to quantify the variable importance. We used the Breiman-Cutler measure of importance (or permutation) measure, the most frequently used measure for random forests. Since each tree is a random subset of the original dataset, the remaining 30% of the

data not selected (e.g., out of bag observations (oob)) can be used to calculate the variable importance. The oob data is used to create permutation accuracy variable importance measure (VIMP) by predicting class membership in the oob sample and then permuting the variables and calculating the predictive accuracy with permuted variables. The average difference in accuracy of the oob versus permuted oob observations over the trees is the VIMP, with a VIMP close to zero implying that the variable has no predictive power. Correlation between variables was assessed to assure that none of the variables were highly ($r>0.75$) predictive of the variables in the model.

Cox Proportional Hazard Regression

Cox regression with all variables were also run for each age and stage model. These models only included main effects and, while the interpretation between RF and Cox PH models differs, were used to assess the benefit of using an RF approach. We found that there were substantive differences in interpretation between the RF and Cox PH mod

Supplementary Table 1. Variable Importance Measures for Age 50 – 54 at Time of Diagnosis

Stage 1/2		
Variable	VIMP	Category
State	0.0064	Macro
Diagnosis Year	0.0053	Macro
Gleason Score	0.1099	Tumor
PSA	0.0575	Tumor
Median Family Income	0.0109	Social
Number of Radiation/Oncologists	0.0032	Access to health
Rural/Urban	0.0019	Social
Number of Chemotherapy Treatment Centers	0.0018	Access to Health
Number of Doctors	0.0013	Access to Health
Social Vulnerability	0.0008	Social
Number of Urologists	0.0008	Access to Health
Medicare/Medicaid Dual Enrollment	0.0004	Social
Race/Ethnicity	0.0002	Race/Ethnicity
Stage 3		
Diagnosis Year	0.0021	Macro
Gleason Score	0.1335	Tumor
PSA	0.0073	Tumor
Race/Ethnicity	0.0055	Race/Ethnicity
Social Vulnerability	0.0038	Social
Number of Doctors	0.0037	Access to Health
Number of Chemotherapy Treatment Centers	0.0018	Access to Health
Median Family Income	0.0018	Social
Number of Radiation/Oncologists	0.0005	Access to health
Prevention Quality Index	0.0004	Access to Health
Stage 4		
Gleason Score	0.0886	Tumor
PSA	0.0582	Tumor
Number of Radiation/Oncologists	0.0053	Access to health
Number of Doctors	0.0045	Access to health
Median Family Income	0.0038	Social
Number of Chemotherapy Treatment Centers	0.0027	Access to health
Prevention Quality Index - African American	0.0018	Access to health
Prevention Quality Index	0.0017	Access to health
Rural/Urban	0.0013	Social
Number of Urologists	0.0013	Access to health
Race/Ethnicity	0.0007	Race/Ethnicity
Social Vulnerability	0.0004	Social

Supplementary Table 2. Variable Importance Measures for Age 50 – 54 at Time of Diagnosis: African American Subsample

Stage 1/2		
Label	VIMP	Category
State	0.0002	Macro
Gleason Score	0.1034	Tumor
PSA	0.0496	Tumor
Median Family Income	0.0045	Social
Number of Doctors	0.0029	Access to Health
Number of Radiation/Oncologists	0.0023	Access to Health
Medicare/Medicaid Dual Enrollment	0.0009	Social
Prevention Quality Index - African American	0.0008	Access to Health
Social Vulnerability	0.0007	Social
Stage 3		
State	0.0156	Macro
Diagnosis Year	0.0021	Macro
Gleason Score	0.1162	Tumor
PSA	0.0248	Tumor
Number of Radiation/Oncologists	0.0075	Access to Health
Race/Ethnicity	0.0022	Race/Ethnicity
Number of Chemotherapy Treatment Centers	0.0019	Access to Health
Prevention Quality Index - African American	0.0015	Access to Health
Prevention Quality Index	0.0013	Access to Health
Stage 4		
GleasonCat	0.0928	Tumor
PSA	0.0537	Tumor
Rural/Urban	0.0013	Social
Prevention Quality Index	0.0013	Access to Health
Number of Chemotherapy Treatment Centers	0.001	Access to Health
Number of Urologists	0.0008	Access to Health
Median Family Income	0.0007	Social

Supplementary Table 3. Variable Importance Measures for Age 55 – 69 at Time of Diagnosis

Stage 1/2		
Variable	VIMP	Category
Diagnosis Year	0.0074	Macro
State	0.0039	Macro
Gleason Score	0.0846	Tumor
PSA	0.0466	Tumor
Race/Ethnicity	0.004	Race/Ethnicity
Medicare/Medicaid Dual Enrollment	0.0012	Social
Rural/Urban	0.0012	Social
Prevention Quality Index	0.001	Access to health
Social Vulnerability	0.001	Social
Median Family Income	0.0009	Social
Number of Urologists	0.0006	Access to health
Number of Radiation/Oncologists	0.0005	Access to health
Number of Chemotherapy Treatment Centers	0.0005	Access to health
Number of Doctors	0.0005	Access to health
Prevention Quality Index - African American	0.0001	Access to health
Stage 3		
Diagnosis Year	0.0066	Macro
State	0.0057	Macro
Gleason Score	0.1153	Tumor
PSA	0.0227	Tumor
Median Family Income	0.0033	Social
Number of Doctors	0.002	Access to health
Medicare/Medicaid Dual Enrollment	0.0016	Social
Social Vulnerability	0.0016	Social
Number of Radiation/Oncologists	0.0013	Access to health
Prevention Quality Index - African American	0.0011	Access to health
Rural/Urban	0.0006	Social
Prevention Quality Index	0.0003	Access to health
Stage 4		
Diagnosis Year	0.0062	Macro
State	0.0046	Macro
Gleason Score	0.0996	Tumor
PSA	0.0577	Tumor
Median Family Income	0.0051	Social
Number of Doctors	0.0031	Access to health
Race/Ethnicity	0.0026	Race/Ethnicity

Number of Radiation/Oncologists	0.0022	Access to health
Rural/Urban	0.0012	Social
Social Vulnerability	0.0011	Social
Number of Chemotherapy Treatment Centers	0.001	Access to health
Number of Urologists	0.0008	Access to health
Prevention Quality Index - African American	0.0006	Access to health
Prevention Quality Index	0.0004	Access to health
Medicare/Medicaid Dual Enrollment	0.0004	Social

Supplementary Table 4. Variable Importance Measures for Age 55 – 69 at Time of Diagnosis: African American Subsample

Stage 1/2		
Label	VIMP	Category
Diagnosis Year	0.0045	Macro
State	0.004	Macro
Gleason Score	0.0875	Tumor
PSA	0.0639	Tumor
Median Family Income	0.0055	Social
Race/Ethnicity	0.005	Race/Ethnicity
Social Vulnerability	0.004	Social
Prevention Quality Index	0.0038	Access to Health
Prevention Quality Index - African American	0.0029	Access to Health
Number of Radiation/Oncologists	0.0019	Access to Health
Rural/Urban	0.0015	Social
Number of Urologists	0.0009	Access to Health
Number of Doctors	0.0007	Access to Health
Medicare/Medicaid Dual Enrollment	0.0004	Social
Number of Chemotherapy Treatment Centers	0.0001	Access to Health
Stage 3		
Diagnosis Year	0.0141	Macro
State	0.0117	Macro
Gleason Score	0.0749	Tumor
PSA	0.0453	Tumor
Race/Ethnicity	0.0047	Race/Ethnicity
Number of Doctors	0.0015	Access to Health
Median Family Income	0.0013	Social
Social Vulnerability	0.0008	Social
Medicare/Medicaid Dual Enrollment	0.0008	Social
Number of Chemotherapy Treatment Centers	0.0005	Access to Health
Prevention Quality Index - African American	0.0004	Access to Health
Prevention Quality Index	0.0002	Access to Health
Number of Radiation/Oncologists	0.0001	Access to Health
Stage 4		
Diagnosis Year	0.0004	Macro
State	0.0063	Macro
Gleason Score	0.0816	Tumor
PSA	0.0537	Tumor
Number of Doctors	0.003	Access to Health
Median Family Income	0.0025	Social
Number of Radiation/Oncologists	0.0021	Access to Health
Prevention Quality Index	0.0018	Access to Health
Rural/Urban	0.0008	Social

Social Vulnerability	0.0007	Social
Number of Chemotherapy Treatment Centers	0.0005	Access to Health
Prevention Quality Index - African American	0.0003	Access to Health
Number of Urologists	0.0001	Access to Health

Supplementary Table 5. Variable Importance Measures for Age 70+ at Time of Diagnosis

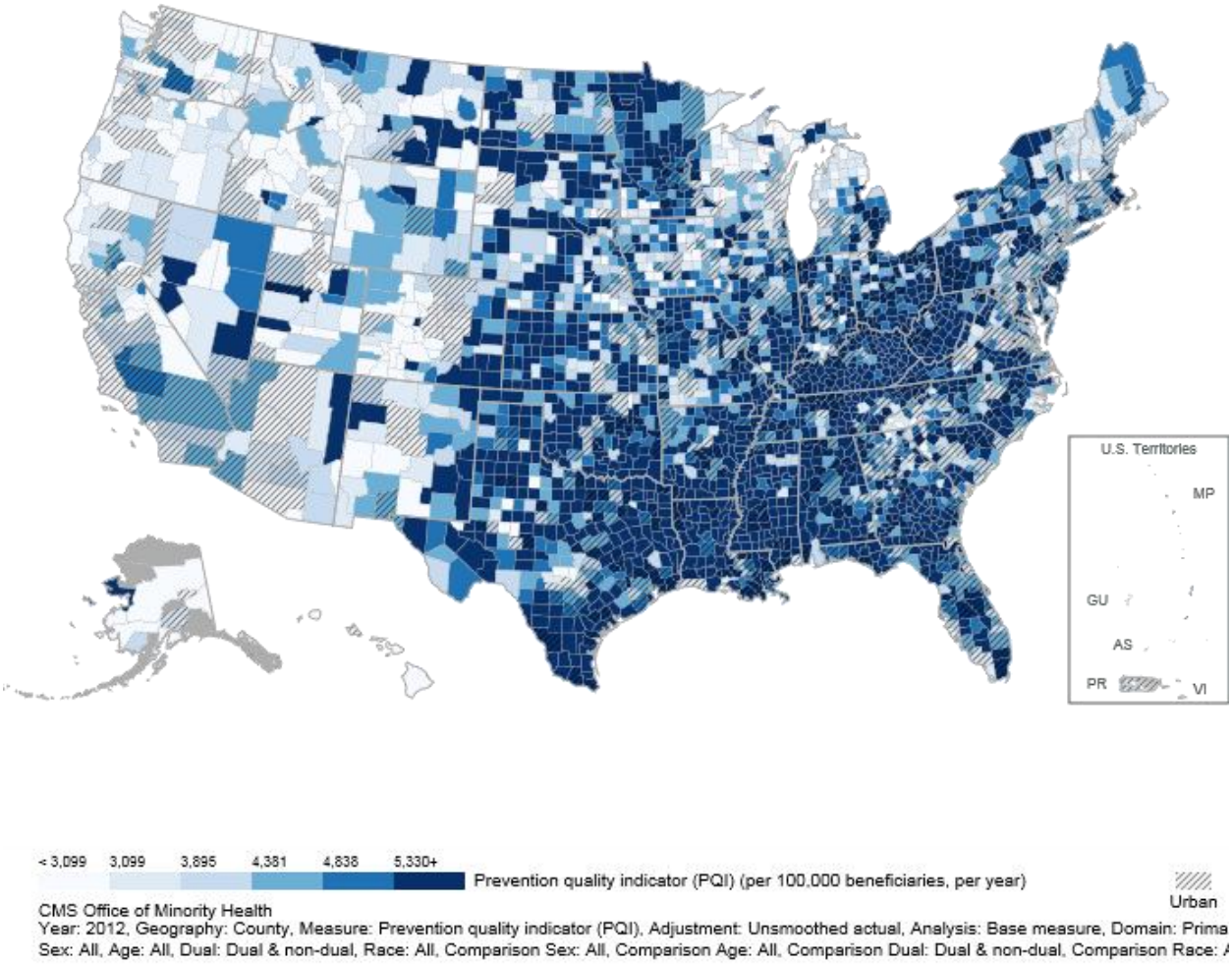
Stage 1/2		
Variable	VIMP	Category
State	0.009	Macro
Diagnosis Year	0.0056	Macro
Gleason Score	0.1106	Tumor
PSA	0.0554	Tumor
Median Family Income	0.0097	Social
Number of Doctors	0.007	Access to Health
Number of Radiation/Oncologists	0.0043	Access to Health
Race/Ethnicity	0.0043	Race/Ethnicity
Social Vulnerability	0.0022	Social
Number of Urologists	0.0021	Access to Health
Number of Chemotherapy Treatment Centers	0.0019	Access to Health
Rural/Urban	0.0019	Social
Prevention Quality Index - African American	0.0011	Access to Health
Medicare/Medicaid Dual Enrollment	0.0009	Social
Prevention Quality Index	0.0007	Access to Health
Stage 3		
State	0.0084	Macro
Diagnosis Year	0.0038	Macro
Gleason Score	0.0947	Tumor
PSA	0.0301	Tumor
Median Family Income	0.007	Social
Race/Ethnicity	0.0059	Race/Ethnicity
Number of Doctors	0.0049	Access to Health
Prevention Quality Index - African American	0.0023	Access to Health
Prevention Quality Index	0.0022	Access to Health
Number of Urologists	0.002	Access to Health
Number of Chemotherapy Treatment Centers	0.0018	Access to Health
Rural/Urban	0.0013	Social
Social Vulnerability	0.0012	Social
Number of Radiation/Oncologists	0.0007	Access to Health
Medicare/Medicaid Dual Enrollment	0.0003	Social
Stage 4		
State	0.0068	Macro
Diagnosis Year	0.0031	Macro
GleasonCat	0.086	Tumor
PSA	0.0209	Tumor

Race/Ethnicity	0.0031	Race/Ethnicity
Number of Doctors	0.0027	Access to Health
Median Family Income	0.0027	Social
Number of Radiation/Oncologists	0.0023	Access to Health
Number of Chemotherapy Treatment Centers	0.0012	Access to Health
Prevention Quality Index - African American	0.0012	Access to Health
Social Vulnerability	0.0012	Social
Rural/Urban	0.0011	Social
Number of Urologists	0.0009	Access to Health
Prevention Quality Index	0.0007	Access to Health
Medicare/Medicaid Dual Enrollment	0.0004	Social

Supplementary Table 6. Variable Importance Measures for Age 70+ at Time of Diagnosis: African American Subsample

Stage 1/2		
Label	VIMP	Category
Diagnosis Year	0.0067	Macro
State	0.0059	Macro
Gleason Score	0.0945	Tumor
PSA	0.0715	Tumor
Median Family Income	0.0129	Social
Number of Doctors	0.0059	Access to health
Number of Radiation/Oncologists	0.0052	Access to health
Race	0.0037	Race/Ethnicity
Number of Urologists	0.0029	Access to health
Number of Chemotherapy Treatment Centers	0.0023	Access to health
Prevention Quality Index	0.0021	Access to health
Social Vulnerability	0.0021	Social
Prevention Quality Index - African American	0.0019	Access to health
Rural/Urban	0.0015	Social
Medicare/Medicaid Dual Enrollment	0.0002	Social
Stage 3		
State	0.0106	Macro
Gleason Score	0.1074	Tumor
PSA	0.0366	Tumor
Median Family Income	0.0044	Social
Prevention Quality Index - African American	0.0011	Access to health
Rural/Urban	0.001	Social
Medicare/Medicaid Dual Enrollment	0.0004	Social
Stage 4		
Diagnosis Year	0.0006	Macro
Gleason Score	0.0891	Tumor
PSA	0.0174	Tumor
Median Family Income	0.002	Social
Number of Urologists	0.0011	Access to health
Number of Radiation/Oncologists	0.0007	Access to health
Rural/Urban	0.0005	Social
Number of Doctors	0.0003	Access to health
Race	0.0002	Race/Ethnicity
Number of Chemotherapy Treatment Centers	0.0001	Access to health

Supplementary Figure 1. Prevention Quality Index by County 2012. *Source CMS Office of Minority Health* https://data.cms.gov/mapping-medicare-disparities.



1. Mérette, C., King, M. C., Ott, J.: Heterogeneity analysis of breast cancer families by using age at onset as a covariate. Am J of Hum Genet, **50:** 515, 1992
2. U.S. Department of Health and Human Services - National Cancer Institute, N.: Cancer Incidence – Surveillance, Epidemiology, and End Results (SEER) Registries Research Data. Bethesda, Maryland: National Cancer Institute, Surveillance Systems Branch, vol. 2017, 2016

3. Epstein, J. I., Egevad, L., Amin, M. B. et al.: The 2014 International Society of Urological Pathology (ISUP) Consensus Conference on Gleason Grading of Prostatic Carcinoma: Definition of Grading Patterns and Proposal for a New Grading System. Am J Surg Pathol, **40:** 244, 2016
4. Yao, N., Foltz, S. M., Odisho, A. Y. et al.: Geographic analysis of urologist density and prostate cancer mortality in the United States. PLoS One, **10:** e0131578, 2015
5. Odisho, A. Y., Cooperberg, M. R., Fradet, V. et al.: Urologist density and county-level urologic cancer mortality. J Clin Oncol, **28:** 2499, 2010
6. Kaiser Commission on Medicaid and the Uninsured and Urban Institute estimates based on data from FY 2008 MSIS. Edited by T. K. F. Foundation: The Kaiser Family Foundation, 2012
7. Breiman, L.: Random Forests. Machine Learning, **45:** 5, 2001
8. Ishwaran, H., Kogalur, U. B., Blackstone, E. H. et al.: Random survival forests. Ann. Appl. Stat., **2:** 841, 2008
9. Archer, K. J., Kimes, R. V.: Empirical characterization of random forest variable importance measures. Computational Statistics & Data Analysis, **52:** 2249, 2008