Survival Analysis of Males in the U.S. with Prostate Cancer Based on Race and Ethnicity

Ranni Tewfik

2024-05-03

# 1. Introduction

Prostate cancer is the second-most common cancer among men in the U.S. and has accounted for more than 34,000 deaths in 2023 from 288,000 new cases.(*Prostate Cancer Prostate Cancer Information and Overview*, n.d.) Other than lung cancer, prostate cancer is the leading cause of cancer death in men – nearly 1 man in 40 will die of prostate cancer.(*Prostate Cancer Prostate Cancer Information and Overview*, n.d.) As prostate cancer is rare in younger men and the risk of prostate cancer increases drastically as men get older (about 60% of cases are in males older than 65 years), age is an apparent risk factor for prostate cancer.(*Prostate Cancer Prostate Cancer Information and Overview*, n.d.) However, the impact of race and ethnicity as a risk factor for death among prostate cancer patients is not completely clear.(*Prostate Cancer Prostate Cancer Information and Overview*, n.d.) According to the American Cancer Society, there is some evidence that African American males have a higher risk than males of other races, but cancer often develops earlier in life in this demographic group.1 Fewer studies have focused on other races and ethnicities, and the reasons for racial and ethnic differences are not well understood.(*Prostate Cancer Prostate Cancer Information and Overview*, n.d.)

A cohort study by Stern et. al (Stern et al., 2021) investigated the association of race and ethnicity with mortality among 51,000 White, Black, and Asian men in Canada with a prostate cancer diagnosis between 1992 and 2010; the study reported that Black men did not have an increased rate of prostate cancer-specific mortality.(Stern et al., 2021) Another cohort study by Rasmussen et al. (Rasmussen et al., 2023) followed 13,000 male veterans with prostate cancer in the U.S. between 2006 and 2021; the study concluded that Hispanic and Non-Hispanic Black males with prostate cancer had lower rates of mortality than their Non-Hispanic White counterparts, and other races with prostate cancer had a higher rate of mortality.(Rasmussen et al., 2023)

Because of the discrepancies in the literature regarding the relationship between race/ethnicity and mortality among patients with prostate cancer, one of the motivations for this project is to better understand that relationship using a large national registry of prostate cancer patients. This project is different from what has already been done because other investigations have used smaller sample sizes or have not included several racial and ethnic groups in the study population. Therefore, the research question for this project is: Among males diagnosed with prostate cancer, are race and ethnicity associated with time to all-cause mortality?

# 2. Methods

## 2.1 Data Source

The Surveillance, Epidemiology, and End Results (SEER) Program of the National Cancer Institute (NCI) provides information on cancer incidence and survival in the United States.(“About the SEER Program,” n.d.) SEER uses population-based cancer registries that cover about half of the national population.(“About the SEER Program,” n.d.) The SEER registries collect data on tumor site and morphology, stage at diagnosis, treatment, vital status, and patient demographics, including race and ethnicity.(“About the SEER Program,” n.d.) This project used the SEER\*Stat software (version 8.4.2) to access the SEER 18 (2000-2018) Registry for data on 1,048,575 prostate cancer patients.

## 2.2 Project Design

This project investigated male patients with prostate cancer in the U.S. from 2000 to 2018. All ages and races/ethnicities were included in the study, however, only patients with localized, regional, or distant cancer stage at diagnosis were included. Patients with unknown/missing survival time were excluded.

## 2.3 Covariates

The main outcome of interest was all-cause mortality. The main exposure of interest was race/ethnicity, which included the following categories: “Non-Hispanic White”, “Non-Hispanic Black”, “Non-Hispanic Asian or Pacific Islander”, “Hispanic (all races)”, and “Other”. The other covariates, age at diagnosis and cancer stage at diagnosis, were categorical variables. The age groups were coded as “0-49 years”, “50-59 years”, “60-69 years”, “70-79 years”, and “≥ 80 years”. The categories for stage at diagnosis were “localized” (i.e., cancer is limited to where it originated), “regional” (i.e., cancer has spread to nearby lymph nodes, tissues, or organs), and “distant” (i.e., cancer has spread to distant body parts).

## 2.4 Statistical Analyses

All analyses were conducted using R software (version 4.3.1). Survival analysis was performed for Kaplan-Meier survival estimates and curves. Cox regression was used to compute adjusted hazard ratios and 95% confidence intervals. The Cox proportional hazards assumption was tested using log-log survival curves and extended Cox models for each covariate. Secondary analyses were conducted for three predictive models: logistic regression, LASSO regression, and discriminant analysis.

# 3. Results

This project included 776,196 males in the U.S. with prostate cancer from the SEER database. Of these, 540,057 (69.6%) were Non-Hispanic White; 526,468 (67.8%) were age 60-79 years at diagnosis; 629,222 (81.1%) had localized stage of prostate cancer at diagnosis; 401,176 (51.7%) had at least a college degree; and 490,476 (63.2%) had a family history of cancer (Table 1). The mean survival time for each ethnic group of prostate cancer patients ranged from 64 months (“Other”) to 78 months (“White”) (Table 1).

*Table 1. Study sample characteristics of males in the U.S. with prostate cancer by race and ethnicity.*

Warning: package 'here' was built under R version 4.3.2

here() starts at C:/Users/ranni/Desktop/MADA/Tewfik-MADA-project



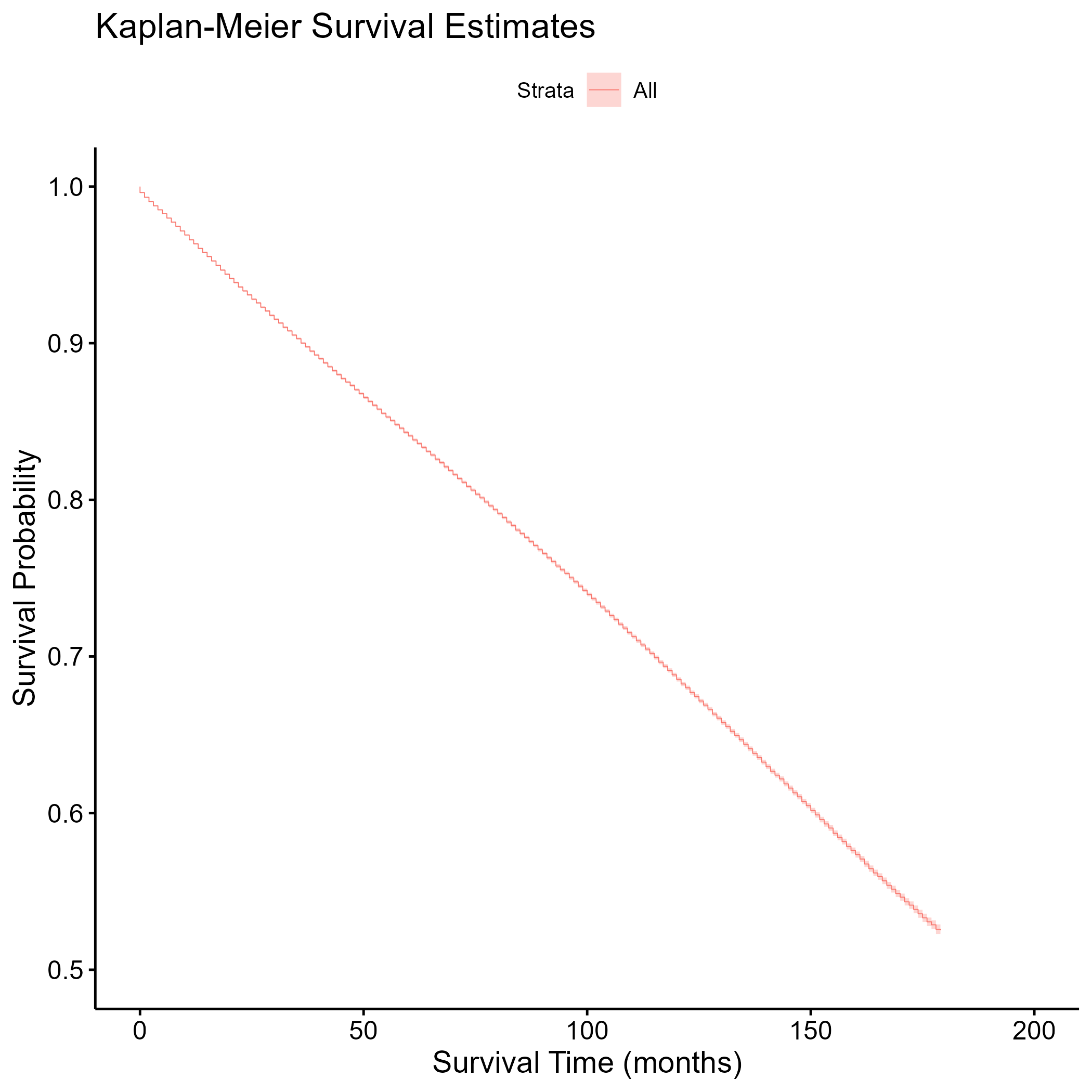
Because the research question for this project is etiologic rather than predictive, the results of the survival analysis approach (etiologic) using Cox proportional hazards models are mainly discussed in this manuscript. However, secondary analyses were also conducted using a non-survival analysis approach (predictive). The results of both analytic approaches are discussed below.

## 3.1 Etiologic Analyses

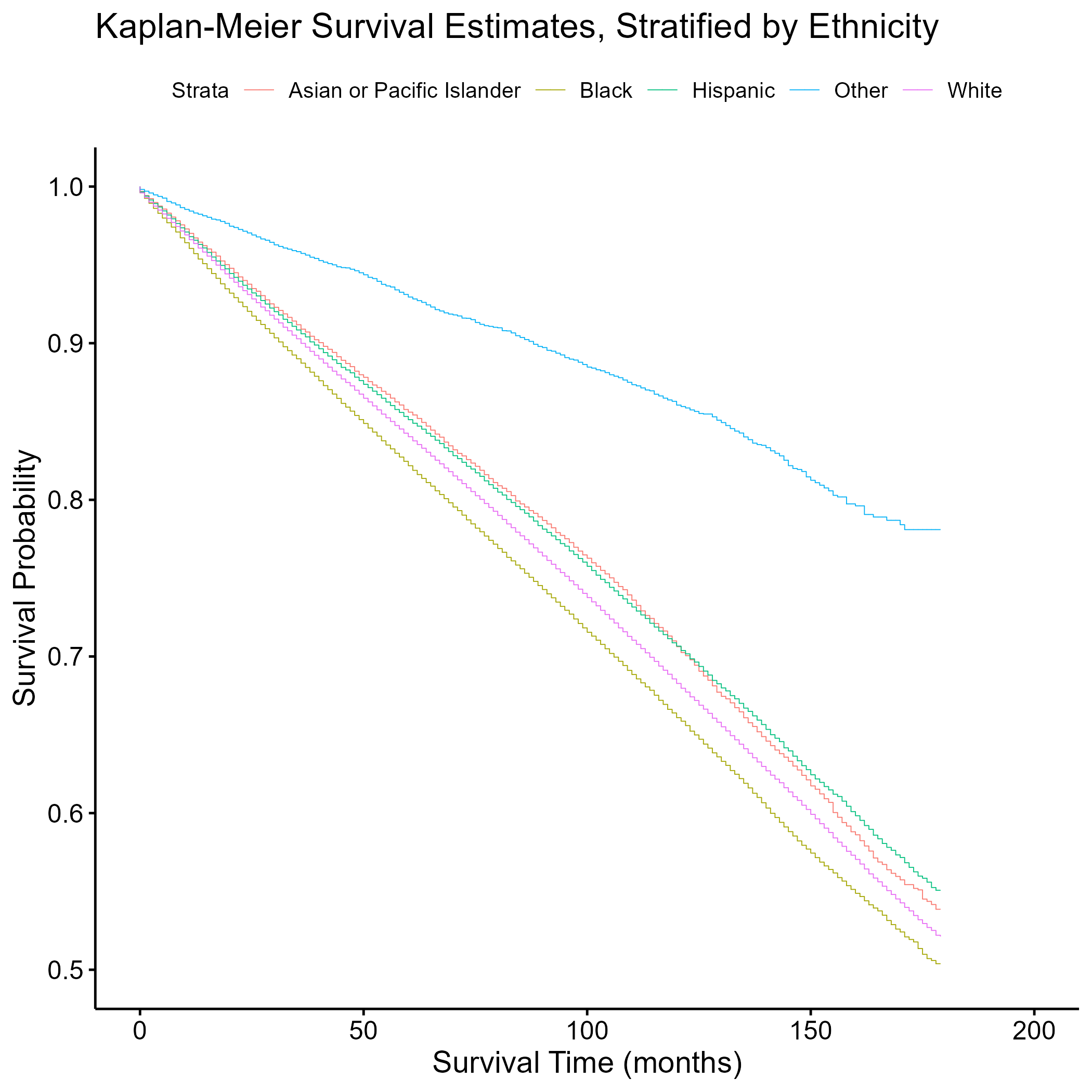
An etiologic (survival analysis) approach was implemented to investigate whether race and ethnicity are associated with time to all-cause mortality. The covariates were age group and cancer stage. Four Cox proportional hazards models were analyzed: three univariable models and the full model with all covariates. Comparing the models with only “ethnicity” (C-index = 0.52), only “agegroup” (C-index = 0.70), and only “stage” (C-index = 0.61), the model with all variables was selected as the best survival analysis model (C-index = 0.76). Backwards elimination analysis supported the selection of the full Cox proportional hazards model.

In assessing time to death (all-cause mortality), the overall 10-year survival was 68.5% (Figure 1). Non-Hispanic Black males had statistically significantly lower survival than other races and ethnicities across the full follow-up time (1-year = 95.7%; 5-year = 82.2%; 10-year = 66.1%) (Figure 2). Based on the log-log survival curves and extended Cox models, the proportional hazards assumption holds for all covariates (see “Supplementary Material” section).

*Figure 1. Kaplan-Meier survival curves for males in the U.S. with prostate cancer.*

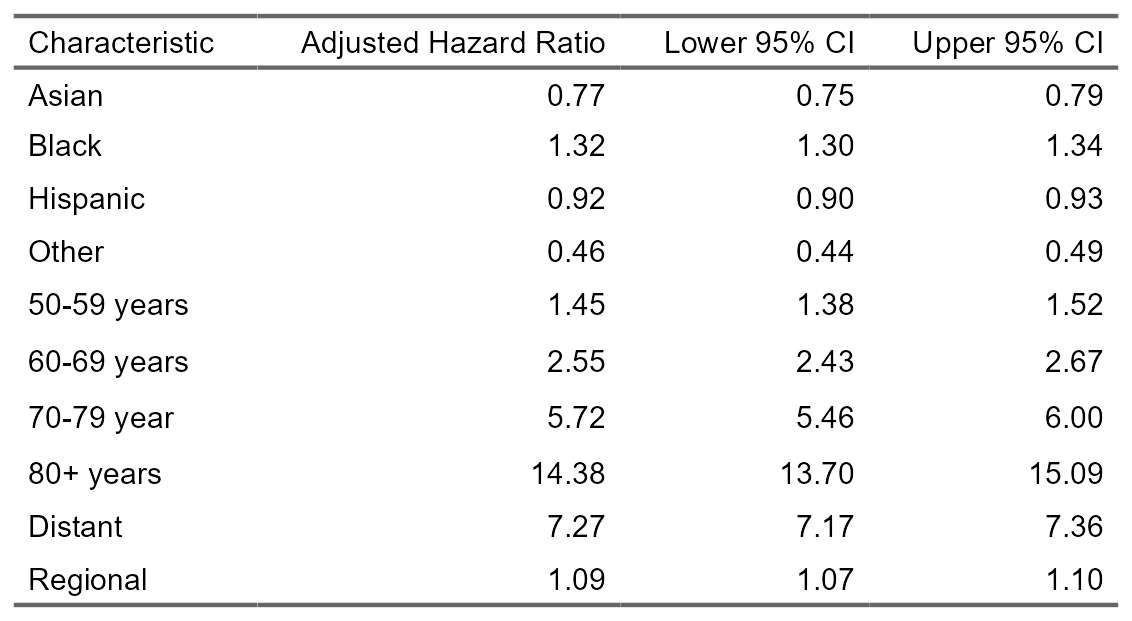


*Figure 2. Kaplan-Meier survival curves for males in the U.S. with prostate cancer, stratified by race and ethnicity.*



After adjustment for age and cancer stage at diagnosis, the rate of all-cause mortality was 1.32 times higher in Non-Hispanic Black males than Non-Hispanic White males (aHR = 1.32; 95% CI = 1.30 – 1.34) (Table 2). Compared to Non-Hispanic White males, the adjusted hazard ratios were less than 1.00 and statistically significant for males who were Non-Hispanic Asian/Pacific Islander, Hispanic, and other race/ethnicity (Table 2). Controlling for the other covariates, males older than 80 years at diagnosis (vs. males 0-49 years) had the highest adjusted hazard ratio for all-cause mortality among all age groups (aHR = 14.38; 95% CI = 13.70 – 15.09), and males with distant stage of prostate cancer (vs. localized stage) had a higher adjusted hazard ratio than males with regional stage (aHR = 7.27; 95% CI = 7.17 – 7.36) (Table 2).

*Table 2. Adjusted hazard ratios and 95% confidence intervals for time to all-cause mortality among males with prostate cancer.*



## 3.2 Predictive Analyses

Because the research question for this project is etiologic rather than predictive, the results of the survival analysis approach using the Cox proportional hazards models are discussed in this manuscript. However, secondary analyses were conducted using a predictive, non-survival analysis approach, which looked at three other full models: logistic regression, LASSO regression, and discriminant analysis. All models were fitted using training data, and the LASSO regression model was tuned using five-fold cross-validation repeated five times to evaluate model performance during the tuning process. The logistic regression model and the discriminant analysis model performed equally better than the LASSO regression model based on the accuracy (0.80) and ROC-AUC (0.76) values. The logistic regression model was chosen as the best predictive model for simplicity of interpretation. A final model evaluation on the logistic regression model using the test data was performed. More details on the predictive analyses are provided in the “statistical-analysis.qmd” file in the GitHub repository.

## 3.3 Supplementary Material

The “Supplementary Material” document includes the results of the exploratory and statistical analyses that were not included in this manuscript. In addition to the Kaplan-Meier survival curves stratified by age group, cancer stage, education, and family history of cancer, the document provides other figures, including the distribution of survival time and stratified plots of ethnicity, survival time, and mortality. Log-log survival curves for ethnicity, age group, and cancer stage are provided as part of the Cox proportional hazards assessment. Finally, tables for the unadjusted hazards ratios and 95% confidence intervals for the effects of ethnicity, age, and cancer stage on all-cause mortality are included.

# 4. Discussion

The results of this project show that compared to Non-Hispanic White males with prostate cancer, Non-Hispanic Black males had the highest rate of all-cause mortality among all races and ethnicities, controlling for age and cancer stage at diagnosis. This finding is in line with the data provided by the American Cancer Society, which reported that Black men 70-110% higher mortality rate for prostate cancer than White men overall in the U.S.(“New Study Shows Western Region has Highest Prostate Cancer Mortality Among White Men in U.S.; Black Men Face Highest Prostate Cancer Mortality Overall,” n.d.) Although the American Cancer Society study (“New Study Shows Western Region has Highest Prostate Cancer Mortality Among White Men in U.S.; Black Men Face Highest Prostate Cancer Mortality Overall,” n.d.) had a higher adjusted hazard ratio than our investigation, the conclusion is the same: Black males with prostate cancer have increased risk of all-cause mortality compared to other racial/ethnic groups. Also, older males and males with more extensive stage of prostate cancer had the highest rates of all-cause mortality, supporting the claim from the American Cancer Society that “your age and overall health… and other factors can also affect your outlook [regarding survival rates for prostate cancer].”(*Prostate Cancer Prostate Cancer Information and Overview*, n.d.) This project establishes an increasing all-cause mortality hazard by age at diagnosis and by cancer stage at diagnosis.

## 4.1 Strengths

The strengths of this project are its large sample size, utilizing data from a national registry, as well as its inclusion of all major racial and ethnic groups in the U.S. Based on the result of the investigation, future research should focus on early diagnosis and treatment among Black males with prostate cancer, especially those who are older and with more extensive stage of cancer.

## 4.2 Limitations

One of the limitations of this project is the use of the SEER 18 registry, which only covers about 28% of the U.S. population.(“SEER\*Stat Databases,” n.d.) The study population may not be a representative sample of the target population, and certain sociodemographic groups may be excluded from the study, resulting in possible external validity problems. Also, the SEER 18 registry only provides data from 2000 to 2018. A longer follow-up time may have resulted in more accurate estimates of hazard ratios. This project examined all-cause mortality as the outcome of interest rather than cause-specific mortality, which provides a smaller sample size (i.e., small changes between comparison groups would not always register as statistically significant) and looks at death specifically caused by prostate cancer instead of death from any cause. This outcome may be more relevant to those with prostate cancer. Another constraint is the limited number of covariates, as few variables of interest were available in the SEER 18 registry. The use of more variables that help explain the association between race/ethnicity and mortality in males with prostate cancer, such as socioeconomic status and comorbidities, may have provided less biased effect sizes.

# 5. References

About the SEER Program. (n.d.). Retrieved from <https://seer.cancer.gov/about/overview.html>

New Study Shows Western Region has Highest Prostate Cancer Mortality Among White Men in U.S.; Black Men Face Highest Prostate Cancer Mortality Overall. (n.d.). Retrieved from <https://pressroom.cancer.org/ProstateCancerMortality>

*Prostate Cancer Prostate Cancer Information and Overview*. (n.d.). Retrieved from <https://www.cancer.org/cancer/types/prostate-cancer.html>

Rasmussen, K. M., Patil, V., Li, C., Yong, C., Appukkuttan, S., Grossman, J. P., … Halwani, A. S. (2023). Survival Outcomes by Race and Ethnicity in Veterans With Nonmetastatic Castration-Resistant Prostate Cancer. *JAMA Network Open*, *6*(10), e2337272–e2337272. <https://doi.org/10.1001/jamanetworkopen.2023.37272>

SEER\*Stat Databases: November 2020 Submission. (n.d.). Retrieved from <https://seer.cancer.gov/data-software/documentation/seerstat/nov2020/index.html>

Stern, N., Ly, T. L., Welk, B., Chin, J., Ballucci, D., Haan, M., & Power, N. (2021). Association of Race and Ethnicity With Prostate Cancer–Specific Mortality in Canada. *JAMA Network Open*, *4*(12), e2136364. <https://doi.org/10.1001/jamanetworkopen.2021.36364>