P8110: Applied Regression II
Homework #4 [15 points]

1. Read the NEJM paper "Long-term survival of participants in the prostate cancer prevention trial" and answer the following questions.

   (a) List three statistical analysis topics that are described in the paper and covered in our class. Copy the sentences used to describe each topic here. [3 points]

   - e.g. 10-year survival rate: "Ten-year survival rates were 83.0% in the finasteride group and 80.9% in the placebo group for men with low-grade prostate cancer and 73.0% and 73.6%, respectively, for those with high-grade prostate cancer."

   1. ***Survival rate****: 15-year survival rate was 78.0% in the finasteride group and 78.2% in the placebo group (similar statement made regarding 10-year survival rate under Results)*
   2. ***Hazard ratio****: the unadjusted hazard ratio for death in the finasteride group was 1.02 (95% CI, 0.97 to 1.08; P=0.46); after adjustment for age, race, and a diagnosis of prostate cancer, the hazard ratio was 1.03 (95% CI, 0.98 to 1.09; P=0.26) (similar statements made about hazard ratios of death after diagnosis of prostate cancer under Results section to summarize Table 2) and also: a hazard ratio of less than 1 indicates a reduced risk of death (Table 2 footnote)*
   3. ***Cox model****: A Cox model was used to estimate hazard ratios for death associated with the two treatments, after adjustment for risk factors. (One model for overall survival, another model for survival after a prostate cancer diagnosis.)*
   4. ***Kaplan-Meier curve****: Figure 2 uses K-M curve to show the overall survival; Figure 3 shows overall survival for men with prostate cancer, stratified by treatment and cancer grade.*
   5. ***Proportional Hazard assumption****: The proportional-hazards assumption for both models was tested within the PROC PHREG module.*
   6. ***Interaction****: The covariates that were included in the model for survival after a prostate-cancer diagnosis were the age at diagnosis as a continuous variable, race (black vs. nonblack), a family history of prostate cancer (yes vs. no), the Gleason score (2 to 6 vs. 7 to 10), and a term for the interaction between treatment and Gleason score.*

7. ***Time-dependent variable****: Diagnosis of prostate cancer was added as a time-dependent covariate to account for the timing of the diagnosis (Table 2 footnote)*
8. ***End point****: The primary end point of the PCPT was histologically confirmed prostate cancer, including prostate cancer that was detected during the course of the 7-year treatment period and prostate cancer that was detected by means of an end-of-study prostate biopsy, as reported previously. (Methods Section p.605)*
9. ***Data censoring****: Data for all men who were last known to be alive or who had died after October 31, 2011, were censored on October 31, 2011. Data for men without a Social Security number were censored on the last date they were known to be alive, according to PCPT follow-up data. (Methods Section p.605)*
10. ***Relative risk*** *: the relative risk of prostate cancer in the finasteride group in current study, as compared with the placebo group, is 0.70; 95% confidence interval [CI], 0.65 to 0.76; $P < 0.001$ (similar statements made several times in similar fashion under Results to summarize Table 1)*
11. ***Median****: the median age of participants was 63.2 years (similar statements made several times in similar fashion under Results to summarize participant demographics)*
12. *Other potential topics mentioned in the paper that did not have much explanations attached:*
    - *randomization*
    - *post hoc analysis*
    - *at-risk analysis*
    - *two-sidedness*

(b) Comment on your answers in question (a). Was the provided description sufficient? Suggest improvement if not sufficient? [3 points]

- e.g. The description of the 10-year survival rate is clear, but can be improved by also providing 95% CI.

1. *Survival rate: Clear description. It could benefit from the inclusion of 95% confidence intervals for their measures.*
2. *Hazard ratio: The adjusted and unadjusted hazard ratio for death in the finasteride group is clear and includes the 95%CI and the p-value.*
3. *Cox model: Clearly specified the covariates in each model. In the model for survival after a prostate-cancer diagnosis, the treatment*

after diagnosis is an important variable as well, but not included in the model.

4. *Kaplan-Meier curve: Using K-M curves is a good way to visualize the survival data. It will be more clear if the y axis is re-sized.*
5. *Proportional Hazard assumption: The paper states that The proportional-hazards assumption for both models was tested, but did not provide the results of the test. It can be more clear with the test results.*
6. *Interaction: Clear description of the interaction in the model and corresponding p-value.*

10. *Relative risk: very clear and contains CI.*
11. *Median age: It could be helpful to add the 95% confidence interval of median age. (Median age should be median survival time)*