Final Project

Caroline Ledbetter

Study Background

The data for this project comes from the Surveillance, Epidemiology and End Results (SEER) Program of the National Cancer Institute. SEER collects information on incidence, survival, and prevalence from specific geographic areas representing 26 percent of the US population. This study specifically explores Lymphoma, a type of cancer in the white blood cells. This study aims to explore differences in survival of patients treated with radiation and surgery vs radiation only.

Methods and Results

Methods Data were collected from three SEER regions (1, 2, and 20). Information on birth year, age, sex, grade of lymphoma (one to nine, with higher grades indicating more aggressive lymphoma), site (Hodgkin or non-Hodgkin), race (white or black), and SEER stage (localized, regional, distant, or unknown) were collected. Individuals who reported they did not receive radiation therapy after cancer diagnosis were excluded. Follow up time in days, survival status at the end of follow-up and cause of death were also recorded. Treatment was defined as radiation with surgery, controls were defined as radiation without surgery.

A propensity score for treatment was calculated using all available covariates. Propensity scores were calculated using logistic regression. An inverse probability weight was calculated for each subject, $w_i = \frac{Z_i}{e_i} + \frac{1-Z_i}{1-e_i}$, where $Z_i =$ an indicator variable denoting whether the or not

the ith subject was treated and e_i is the propensity score.[1] The inverse probability weight was then used to obtain balanced treatment and control groups. Table 1 shows the participant characteristics by treatment group with and with out weighting. To determine if covariates varied across treatment groups, categorical variables were tested using χ^2 and continuous variables were tested using ANOVA. For the weighted analysis, continuous variables were tested using logistic regression. Inverse probability of treatment weighting using the propensity score allows us to obtain an unbiased marginal treatment effect or an estimate of the average treatment effect using cox hazard regression[1].

Cox hazard regression was used to obtain a hazard ratio for all-cause mortality and death from lymphoma. Weights were incorporated using the survey package v(2.0.32)[2] in R version 3.3.2 (2016-10-31)[3]. A sensitivity analysis was performed.

Results The data contained 31,689 observations, 22,505 were excluded because they did not receive radiation treatment, leaving 9,184 for analysis. Participants receiving treatment were more likely to have Hodgkin lymphoma and to be from location 20, treated participants were also more likely to be black and be younger. Inverse probability weighting resulted in balanced treatment groups. (Table 1). The maximum follow-up time was 347 days. Individuals who received treatment were more likely to survive to the end of follow-up (Fig 1). The estimated hazard ratio for all cause mortality was 0.88 (95% CI: 0.82 - 0.95), thus undergoing surgery in addition to radiation reduced the hazard of all cause mortality by 12%. The estimated hazard ratio for death due to lymphoma was 0.85 (95% CI: 0.77 - 0.94), thus undergoing surgery in addition to radiation reduced the hazard of death due to lymphoma by 15%.

Conclusions

References

1. Austin P. The use of propensity score methods with survival or time-to-event outcome: Reporting measures of effect similar to those used in randomized experiments. *Statistics in Medicine*. 2014;33:1242-1258.

2. Lumley T. Analysis of complex survey samples. *J of Statistical Software*. 2004;9(1):1-19.

3. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2016. https://www.R-project.org/.