Analyzing Previous Results of Clinical Trials to Determine the Ethical Use of Medication in Late Stage HCC Patients

We are analyzing data given about outcomes and risk factors of patients in previous clinical trials conducted with the drug in question, to determine if the outcomes were sufficient enough to justify the use of the drug to treat OHSU’s late stage HCC patients. The data about the mortality of these patients is categorical with two categories, survival or death. Therefore, tests that use a categorical response variable are conducted.

First the chi-squared test is used, as this analyzes independent categorical data.

We use chi-squared test with all categorical variables against survival and find evidence that there is a correlation between a patient’s chance of survival and their lack of symptoms (= 15.524, p-value = 8.144e-05), as well as a correlation between metastatic cancer and mortality (= 12.654, p-value = 0.0003748). Therefore, we believe that should OHSU decide to move forward with participation in the clinical trial, it’s participants should only include those that show symptoms, and those whose tumors have not metastasized.

Some points of data included are continuous, limiting the use of chi-squared tests. Logistic regression can be used to determine a relationship between a categorical response variable and either continuous or categorical variables (Le, 2018), and reveals that there is a significant relationship between survival and age (p-value = 0.008482). A frequency table of survival and age (binned for every decade 20 to 100 years; 8 bins total) shows that the rate of survival is lower for patients between the ages of 60 and 90 (44.9% survived), than it is for patients between ages 20 to 60 (61.4% survived).

Interestingly, a correlation is not found when gender, AFP level, number of nodules, size of largest nodule, and lifestyle factors such as alcohol and smoking, are compared to survival (p-value > 0.05). This means that those variables do not significantly impact a patient’s survival, and therefore should not impact a patient’s ability to qualify for the clinical trial.

*Methods:*

We are interested in analyzing the data given regarding the patients’ tumors (quantity and size), their AFP levels, and mortality to determine the efficacy of the drug and whether or not treating our late-stage HCC patients with the medication is an ethical choice. The first method chosen is the chi-squared test as it tests independent, categorical data. For this method, the p-value is determined and compared to a significance level of 0.05. In this scenario, our primary interest is in patient survival as a quantitative metric for success of the drug. Chi-squared tests are primarily used, as we are determining if relationships existed between two categorical variables, such as symptoms and survival. The second method to be used is logistic regression. Logistic regression can be used to determine relationships between categorical response variables, and either continuous or categorical variables. We are using this method to determine the relationship between age and mortality in patients on the drug in the clinical trial.

Frequency tables are used to further explore the data when the above methods show statistically significant correlations between survival and other variables. It is used to further explore how the variables, specifically categorical variables, interact with each other. Conclusions will be drawn based on correlations found between the various variables, and survival.

*Results:*

The test methods are tested in the statistical programming language R. Our results indicate that patients who are asymptomatic or are over 60, do not benefit from the clinical trial drug, as the survival rate is not affected at a statistically significant level. Therefore, as research doctor and head of the patient coalition, we feel that participation in the drug’s clinical trial could be beneficial for patients under the age of 60 who show symptoms of HCC. Because of the reported history of negative side effects from drugs like the one in question, it is our recommendation that those patients who do not fit the above criteria, should not be given the drug, as there is a high chance that their quality of life would be impacted negatively.

*References:*

Le, James. (2018, April 10). *Logistic Regression in R Tutorial* Datacamp.

<https://www.datacamp.com/community/tutorials/logistic-regression-R>

*R Code:*

# Setting up the R workspace -----------------------------------

# import the data set

hcc = read.csv('HCCdata.csv')

#chi-squared test on all categorical variables against Survival

chisq.test(table(hcc$Survival, hcc$Symptoms))

chisq.test(table(hcc$Survival, hcc$AFP))

chisq.test(table(hcc$Survival, hcc$HepB))

chisq.test(table(hcc$Survival, hcc$HepC))

chisq.test(table(hcc$Survival, hcc$Metastasis))

chisq.test(table(hcc$Survival, hcc$Gender))

#logistic regression on all variables with Survival as the response variable

glm.all = glm(Survival~ ., data = hcc, family = "binomial")

summary(glm.all)

#frequency table for survival and age

bins = seq(20, 100, by=10)

x = cut(hcc$Age, bins)

table(hcc$Survival, x)