Treatment With Indinavir Results in a Reduced Hazard of AIDS Diagnosis or Death: Cox Proportional Hazards Model and Power Analysis of Hammer et al., 1997

Carrie Saada

April 8, 2019

```
knitr::opts_chunk$set(message=FALSE, warning=FALSE, fig.height=4, fig.width=5,
                      fig.align = "center")
library(tidyverse)
library(broom)
library(survival)
library(survminer)
library(coxed)
```

```
AD = read.csv("AIDSdata.csv")
```

Introduction

Combinations of different drugs in treatment are an important aspect of HIV therapy to prevent the virus developing resistance to any one drug. For this reason, development of new drugs is vital. Hammer et al. performed a controlled trial using a protease inhibitor called indinavir (IDV) in combination with two nucleotide analogues (Hammer et al. 1997). In this paper, I analyze the data collected by Hammer et al using a Cox Proportional Hazards model to investigate the effect of indinavir on survival. Additionally, I analyze the power of the Hammer study using a simulation of survival data with similar parameters. I hypothesize that indinavir will have a lower hazard of death or AIDS diagonsis than the control group, and that the Hammer study will have a moderately high power (high due to the large sample size, and moderately so due to the low number of events).

Methods

Cox Proportional Hazards Model

I developed a Cox Proportional Hazards model with multiple covariates to fit the data. The response variable of interst was time to AIDS defining diagnosis or death.

I chose the explanatory variables for the model using backward selection. I began with a fully interactive model with the explanatory variables: whether the patient was treated with IDV (tx), the patient's CD4 stratum at screening (strat4), their sex (sex), race/ethnicity (raceth), IV drug use history (ivdrug), whether the patient was a hemophiliac (hemophil), their score on the Karnofsky Performance Scale (karnof), their baseline CD4 count (CD4), the number of months of prior ZDV use (priorzdv), and their age at enrollment (age). I eliminated variables based starting with those that had the highest p-value in their interaction terms, then checked that the resulting Cox PH model was equivalent to the previous model using the likelihood ratio test. I eleminated the following variables in order: whether the patient was a hemophiliac, the patient's CD4 stratum at screening, all interaction terms with a p-value greater than 0.1, all interaction terms with the patient's Karnofsky score (the term for the score without interaction remained in the model), and the interaction term for race/ethnicity and age. I then checked the resulting model against the original full

model using the likelihood ratio test to double check the result. Because treatment with IDV had a high (>0.05) p-value in the resulting model, but its interaction terms were less than 0.05, I compared the model I developed to a model without the treatment with IDV variable or its interaction terms. I analyzed the effect of including the IDV treatment variable using the likelihood ratio test between the two models.

Power Analysis

Power analysis is valuable because it allows us to plan the sample size of experiments based on an estimate for the minimum effect size we expect. Most experiments aim for a power of at least 0.80 (Whitlock and Schluter 2015). That is, in a situation where the null hypothesis is false, we will successfully reject the null hypothesis 80% of the time. I investigated the effect of sample size and effect size on simulated survival data with similar parameters to the Hammer et al study.

I used sim.survdata from the coxed R library to simulate the data. The Hammer et al data used for this project had 69 events out of 851 patients, so I used 782/851 for the proportion of censored observations in the simulated data. Sample size and effect size were varied as described below. sim.survdata uses a normal distribution for the explanatory variable, rather than a binary categorical variable. I did not include the patient-specific variables that were not controlled by the study. For each value of the varied parameter, I simulated 1,000 data sets and calculated the proportion that were significant ($p < 0.05$).

Sample Size

To determine the effect of sample size on power, I used the previously described technique with an effect size of beta = -1.454. This is the effect size I determined in my Cox PH model for the data. I ran the simulation for sample sizes of 100, 400, 800, 851, 1200, 1600, and 4000. I used 851 entries of Hammer et al's data (a small number of entries were not used to remove a treatment group irrelevant to this analysis), so I used this number to represent the sample size of the study.

Effect Size

To determine the effect of effect size on power, I used the same technique as before but varied the effect size, keeping sample size constant at 851, the size of our data set. I set beta equal to 0, -0.2, -0.5, -1, -1.454, -1.5, -2, and -3. Note that beta = 0 is a null sample, and that beta = -1.454 is the value calculated for that coefficient by my Cox PH model of the Hammer et al data.

Results

Cox Proportional Hazards Model

I developed I Cox Proportional Hazards model for the event of diagnosis with AIDS or death, as explained by treatment with IDV and a series of patient-specific factors.

```
## # A tibble: 11 x 7
##    term           estimate std.error statistic   p.value conf.low conf.high
##    <chr>             <dbl>     <dbl>     <dbl>     <dbl>    <dbl>     <dbl>
##  1 tx              -1.45     0.863       -1.69   9.19e-2 -3.14      0.237
##  2 sex              2.47     0.867        2.85   4.39e-3  0.770     4.17
##  3 priorzdv        -0.0153   0.0104      -1.47   1.41e-1 -0.0358    0.00507
##  4 raceth           1.67     0.625        2.67   7.55e-3  0.445     2.90
##  5 ivdrug          -0.855    0.337       -2.54   1.12e-2 -1.52     -0.195
##  6 karnof          -0.0563   0.0146      -3.86   1.15e-4 -0.0849   -0.0277
```

```
##  7 cd4             -0.0153   0.00315     -4.85   1.21e-6 -0.0214   -0.00910
##  8 tx:sex           1.39     0.666        2.09   3.67e-2  0.0861    2.70
##  9 tx:priorzdv     -0.0346   0.0142      -2.44   1.47e-2 -0.0625   -0.00682
## 10 sex:raceth      -1.60     0.578       -2.77   5.65e-3 -2.73     -0.467
## 11 priorzdv:ivdr~   0.0174   0.00670      2.60   9.39e-3  0.00427   0.0305
```

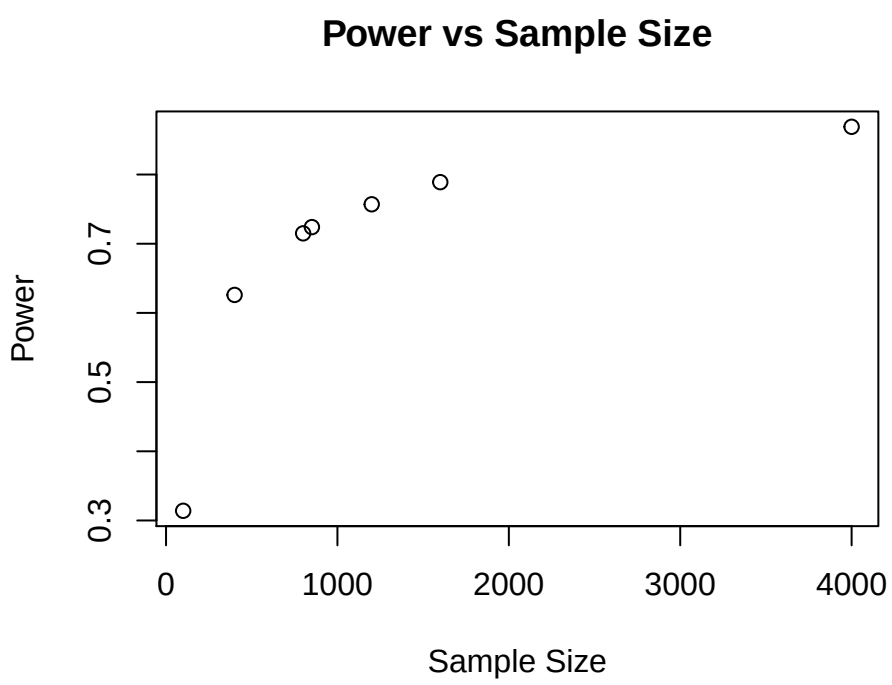
The coefficient for treatment with IDV is -1.454 (95% confidnece interval [-3.145, 0.237]). This tells us that the hazard ratio between the treatment and control groups is $e^{-1.454} = 0.234$, meaning that the treatment group has a lower hazard than the control group. However, the 95% confidence interval includes zero, so it is possible the opposite is true. The treatment variable interacts with both the patient's sex and any prior treatment with ZDV, and these variables are significant to the model with p-values of less than 0.05. To asses the significance of the treatment group to the model, I compared the likelihood of a model with the treatment variable and interactions to the likelihood of a model without either. I found that the p-value for that likelihood ratio test was 0.0003078, which is less than 0.05 by two orders of magnitude. Based on this analysis, the treatment with IDV does have a significant and beneficial effect on patient survival.

Power Analysis

Sample Size

I tested the effect of sample size on the power of the study using 1000 simulated samples per sample size, with a constant beta coefficient of -1.454.

Sample Size	Power
100	0.314
400	0.626
800	0.715
851	0.724
1200	0.757
1600	0.789
4000	0.869

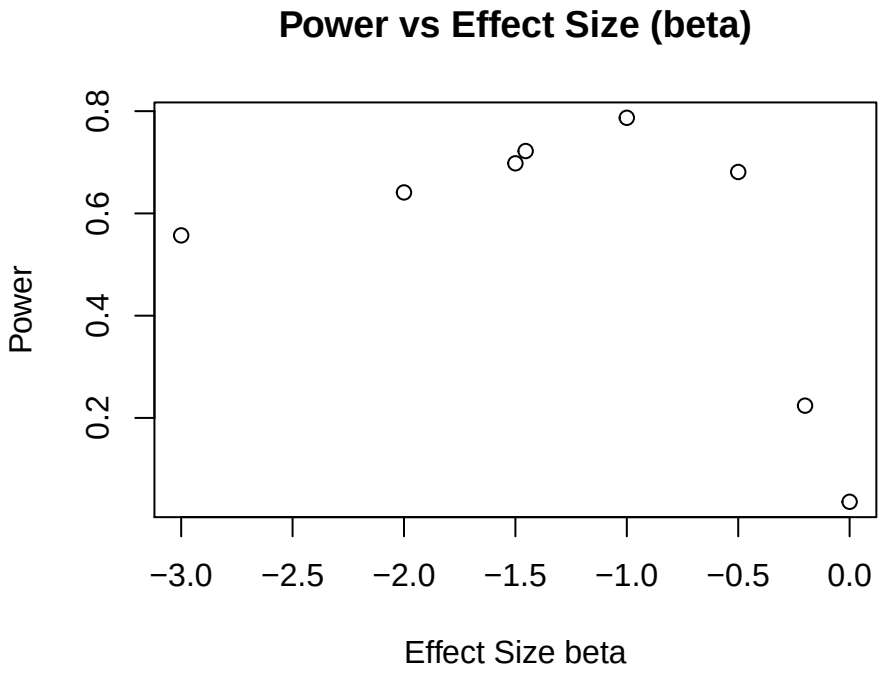


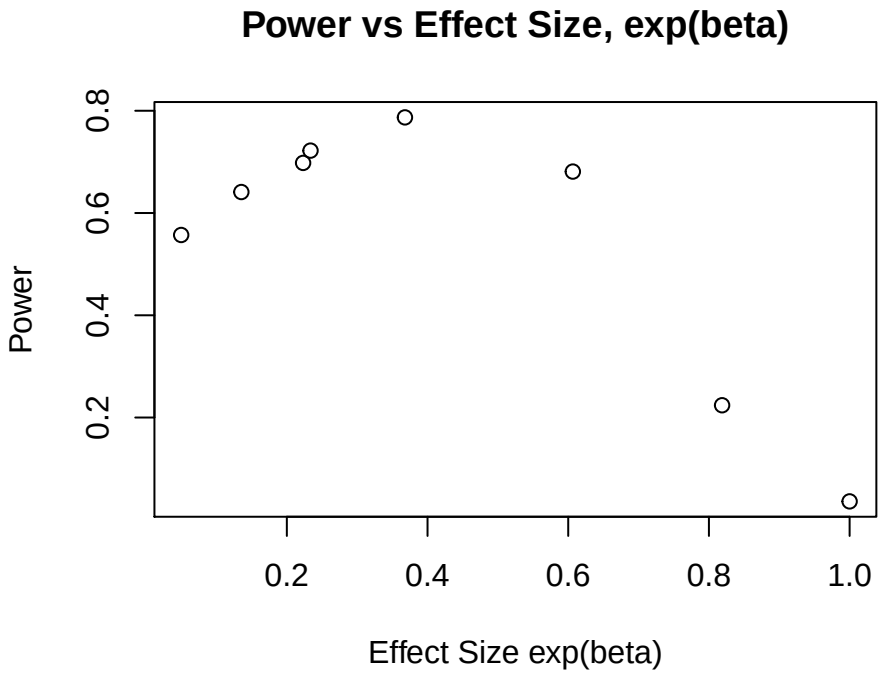
Power increases with sample size, in a concave down manner. Based on this analysis, the power of a Cox Proportional Hazards analysis on a study with 851 subjects, the same proportion of censored obserations, and the same beta coefficient as the Hammer et al study is 0.724.

Effect Size

I tested the effect of effect size on the power of the study with 1000 simulated samples per effect size, with a constant population of 851. The coefficient for a Cox PH model is beta, and the hazard ratio is exp(beta), so I've provided both numerically.

Beta	Exp(Beta)	Power
0	1	0.036
-0.2	0.8187	0.224
-0.5	0.6065	0.681
-1	0.3679	0.787
-1.454	0.2336	0.722
-1.5	0.2231	0.698
-2	0.1353	0.641
-3	0.0498	0.557





Power increases as the magnitude of the effect increases up to a point, then begins to slowly decrease. As the effect size gets farther from zero, for beta, or farther from one, for exp(beta), the power increases. Note that because exp(beta) is a ratio, a larger magnitude of multiplicative difference is distance from 1, not zero. The peak is near beta = -1, or exp(beta) = 0.3679, which has a power of 0.787. As beta continues to get farther

from zero and exp(beta) gets farther from 1 past the peak, the power decreases again.

The power using the effect size calculated using the Cox PH analysis for the Hammer et al data is 0.722, which is very similar to the value of 0.724 from the sample size analysis. Since for those two entries all of the parameters were identical, that consistency is encouraging for the validity of this analysis.

Discussion

Using data from Hammer et al, I found that treatment with IDV significantly improves survival outcomes with a hazard ratio of 0.234 in a manner dependent on interactions with the patient's sex and prior treatment with ZDV. Interacting variables make it difficult to analyze significance, so I compared the model with IDV treatment and interaction to the model without either and found a significant difference in likelihood between the two models. This result is dependent on multiple patient-specific factors, so the full analysis should be used to determine if IDV will increase survival for a specific patient, rather than generalizing from the hazard ratio alone.

I analyzed the effect of sample size and effect size on the power of a Cox PH model with similar parameters to the study by Hammer et al. I determined that power increases with sample size in a concave down manner, and that power increases as the hazard ratio decreases from 1 until approximately 0.3679 and then begins to decrease. The power for the Hammer study was below the threshold of 0.80, but by my analysis would have had to approximately double its sample size to achieve this. However, these power analyses have only limited applicability to the Hammer et al study. The simulated data did not account for patient-specific factors which were significant in the actual data. Additionally, the simulation used a continuous, normally distributed explanatory variable, rather than the binary explanatory variable used by Hammer et al. Power analysis of Cox PH models could be improved by the inclusion of these factors to better match the reality they model.

Bibliography

Cohen, Jacob. "Statistical Power Analysis." Current Directions in Psychological Science, vol. 1, no. 3, June 1992, pp. 98–101. SAGE Journals, doi:10.1111/1467-8721.ep10768783.

Dorey, Frederick J. "In Brief: Statistics in Brief: Statistical Power: What Is It and When Should It Be Used?" Clinical Orthopaedics and Related Research, vol. 469, no. 2, Feb. 2011, pp. 619–20. PubMed Central, doi:10.1007/s11999-010-1435-0.

Hammer, Scott M et al. "A Controlled Trial of Two Nucleoside Analogues plus Indinavir in Persons with Human Immunodeficiency Virus Infection and CD4 Cell Counts of 200 per Cubic Millimeter or Less." NEJM. Sep 1997.

Kuiper, Shonda, and Jeffrey Sklar. Practicing Statistics : Guided Investigations for the Second Course. Pearson, 2013.

Whitlock, Michael, and Dolph Schluter. The Analysis of Biological Data. Second ed., Roberts and Company, 2015.

Appendices

The code used for these analyses can be found in "survival-project-code.Rmd" in the GitHub repository with this project.