Analysis of Survival Functions In Predicting Length Of Stay In Florida Hospitals

Pali Sen
University of North Florida, psen@unf.edu

Benjamin Webster
University of North Florida, ben.webster@nlplogix.com

Journal of Basic and Applied Research International
19(3): 194-205, 2016
ISSN: 2395-3438 (P), ISSN: 2395-3446 (O)

International Knowledge Press
www.ikpress.org

ANALYSIS OF SURVIVAL FUNCTIONS IN PREDICTING LENGTH OF STAY IN FLORIDA HOSPITALS

WEBSTER BENJAMIN[1] AND SEN PALI[1*]
[1]University of North Florida, Jacksonville Florida, 32224, USA.

AUTHORS' CONTRIBUTIONS

This work was carried out in collaboration between both authors. Author SP designed the study, wrote the protocol and interpreted the data. Author WB anchored the field study, gathered the initial data and performed preliminary data analysis. Both authors searched the literature and produced the initial draft and approved the final manuscript.

Received: 10th November 2016
Accepted: 21st December 2016
Published: 23rd December 2016

Original Research Article

ABSTRACT

Statistical methodology and data analytics have avenues of exploring relationships among observed variables that are qualitative and quantitative in nature. The main objective of this study is to show that there is not a single "best" model to predict the length of stay of elderly patients; but rather that there is a preferred model for different age groups with various health conditions. We investigate a large amount of public data that are collected for the Agency for Health Care Administration and suggest possible predictive models to interpret its outcomes. Our data consist of every Medicare inpatient hospital discharge record related in the state of Florida 2011 related to the following primary diagnoses: Acute Myocardial Infarction, Heart Failure, and Pneumonia. The response variable is duration of stay in days. The nature of the predictor variables is either categorical or ordinal. We use an Accelerated Failure Time model and a Cox Proportional Hazard model for the right-censored response time in order to analyze related distribution functions. We interpret the effect of sex, primary diagnosis, age, inclusion of respiratory charges, and severity of illness as explanatory variables and use these to rank the patients in terms of expected length of stay. We use an extensive amount of visual display to substantiate the outcomes. The result includes expected instantaneous rate of change on the hazard functions of Accelerated Failure Time and Cox Proportional Hazard models, as well as the Kaplan Meier estimates. The study results indicate the importance of using multiple model types when analyzing any data which incorporates failure time data.

Keywords: Accelerated failure time model; Cox proportional hazard model; Kaplan Meier estimates; saturated model; reduced model; survival analysis.

1. INTRODUCTION

Over the past few years there has been an increased focus on the quality of treatment given to patients in the hospital and how this affects the likelihood of a patient returning to the hospital for the same condition, specifically within 30 days of discharge. The Affordable Care Act includes many specific rules under the blanket of the Hospital Readmission Reduction Program (HRRP) [1] that penalize hospitals for not providing adequate care to their patient base. Specific focus has been given to three

**Corresponding author: Email: psen@unf.edu;*

distinct primary diagnoses groups: Acute Myocardial Infarction (AMI), Heart Failure (HF), and Pneumonia (PN). The focus is also initially on those patients which are covered by government funding exclusively patients 65 and older on Medicare or Medicare Managed Care. The Affordable Care Act incentivizes the hospitals to better care for their patients in order to reduce readmissions. In order to aid the understanding of the population of interest, we analyze the publicly available Agency for Health Care Administration (AHCA) in Florida [2,3] inpatient dataset for calendar year 2011, which has a record for every inpatient hospitalization during that year, along with the patients' demographic information and the common discharge summary, including length of stay (in days) and medical diagnoses with charges.

Webster and Sen [4] analyzed length of stay data using a loglinear model and identified some important covariates that should be used to better understand the reason for stay duration. They found that the frequency of patients with a specific level of length of stay was highly dependent on the patient's primary diagnosis, sex, and the existence of comorbidities. Their analysis showed, for example, that the most frequent length of stay for HF patients with no comorbidities was 2 days, whereas similar patients with major complicating comorbidities most frequently stayed 4 days. In contrast, we now investigate the conditional probability that a patient will be released on a set day, given that they have not already been discharged prior to that day. The main objective of the study is to show that there is not a single "best" model to predict the length of stay of elderly patients, rather a preferred model for different age groups, with various health conditions.

We compare the survival probabilities generated by a parametric Accelerated Failure Time (AFT) model [5] and a nonparametric Cox Proportional Hazard (CPH) model [6] and discuss similarities and differences presented by the two approaches. The interpretation of the differences in survival times is based on the level of the categorical predictors as well as the age of the patient, which is an ordinal predictor. Of particular interest to the study is the interaction between these variables. It seems reasonable to believe that increased age would lead to increased length of stay, but is this particularly true for one diagnosis type and not evident for another diagnosis type?

2. MATERIALS AND METHODS

The analysis is for Medicare and Medicare Managed Care patients 65 or over with principal diagnosis codes corresponding to AMI, HF, or PN, as identified by the ICD – 9 code [2] with the response variable length of stay, which is measured in days. The impact of complicating conditions, known as comorbidities, is also investigated for three levels of comorbidity: None, Complicating Comorbidities (CC), and Major Complicating Comorbidities (MCC). These values range from 0 days to over 400 days. Only patients released to their homes are considered as data for this analysis. We remove patients who die in the hospital or who are released to hospice. We are interested in modeling and analyzing data whose end point is the time until a discharge occurs. From this perspective we can see the length of stay as failure time data, discharge from the hospital as the event, and model associated probabilities of discharge from the hospital at a given time t [7].

Two methods of failure time modeling are investigated to predict length of stay for patients with diagnoses of interest to HRRP, the bill passed as the part of the Affordable Care Act. Our goal is to determine the accuracy of the predictions on out of sample data on two methods and interpret the influence of explanatory variables on the response.

Failure time data occur when subjects are at risk under different experimental conditions and are a focus of analysis of failure time data is on regression of survival data to estimate regression coefficients and distributional shape with censoring. Ultimately we assess the dependence of failure time on explanatory variables [8]. Explanatory variables considered for this model are

1. Sex – Categorical – {M,F}
2. Age – Ordinal Integer – [65 – 108]
3. Diagnosis – Principal Diagnosis – Categorical – {AMI, HF, PN}
4. Comorbidities – Complicating Conditions – Categorical – {None, CC, MCC}
5. Respiratory Charges – Existence of Respiratory Charges – Indicator Variable – {0, 1}

For this analysis, time, T is a discrete random variable taking values $x_0 < x_1 < \cdots < x_{16} < x_{17+}$, where x_i is the i-ith day after admittance to the hospital. The probability density function

$$f(x_i) = P(T = x_i) \quad (1)$$

has a corresponding survival function. The function associated with this distribution would be interpreted as the probability that T, the time in days until the patient is discharged, is as least as large as the value t. This can be calculated as

$$F(t) = \sum_{j|x_j \geq t} f(x_j) = f(x_t) + f(x_{t+1}) + \cdots + f(x_{17+}) = \sum f(x_j) H(x_j - t) \quad (2)$$

where $H(x)$ is the Heaviside function as defined below.

$$H(x) = \begin{cases} 0, & x < 0 \\ 1, & x \geq 0 \end{cases} \quad (3)$$

The use of the Heaviside Function becomes clear as we try to estimate these functions. Notice $F(t)$ is not an increasing function but it is a monotonically decreasing function and $F(0) = 1$. We then define the hazard function at x_j as the conditional odds of failure at x_j

$$\lambda_j = P(T = x_j | T \geq x_j) = \frac{f(x_j)}{f(x_j)+f(x_{j+1})+\cdots+f(x_{17+})} = \frac{f(x_j)}{F(x_j)} \quad (4)$$

The hazard function could be intuitively understood as an informing probability density function, or more simply, the question that a patient wants to know every morning what the chance of discharge on that particular day happens to be. The baseline hazard function is calculated as

$$\lambda_0 = \frac{f(x_0)}{F(x_0)} = \frac{f(x_0)}{1} = f(x_0), \quad (5)$$

implying that at the moment of admission to the hospital the survival function and hazard function are identical. At each new day the probability that a patient is being discharged on the previous day is 0. The probability that they will be discharged on the present day, and every day beyond that, increases as a result. For finite (censored) discrete data, there is a one-to-one relationship between the probability density function, the survival function, and the hazard function. This relationship is defined as

$$F(t) = \prod_{j|x_j<t} (1 - \lambda_j) \quad (6)$$

and

$$f(x_j) = \lambda_j \prod_1^{j-1} (1 - \lambda_i). \quad (7)$$

3. RESULTS

We investigate the response variable to identify the shape of the distribution. It indicates a Poisson distribution for the length of stay data. Since discharge dates of the patients are the data points for the hospital stay and are provided in whole day increments, we categorize length of stay in seventeen groups and one for higher values. This higher value group is not included in the analysis as it makes a less meaningful contribution than the groups. Hence, the data will be right censored. The right-censored observations occur when length of stay is seventeen days or over and make up 3.35% of the total data.

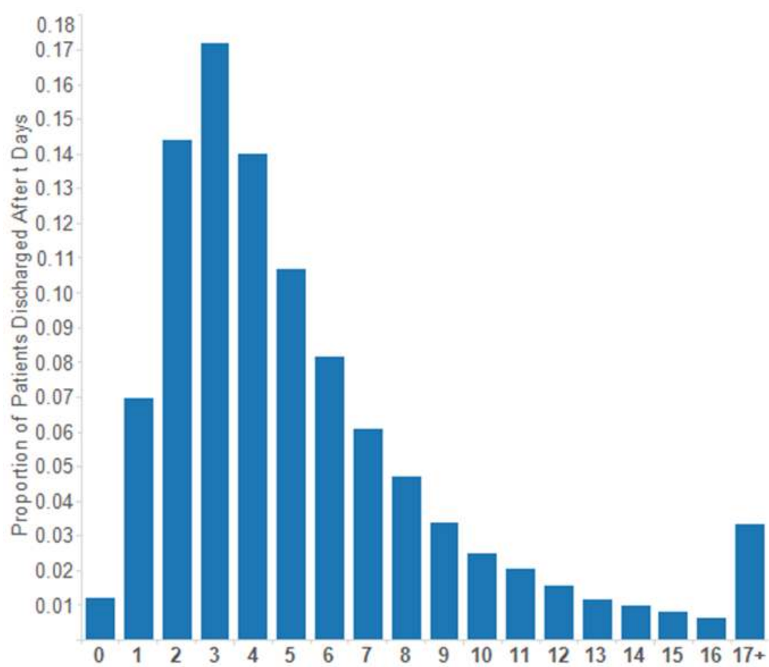

Fig. 1. Distribution of length of stay data for the florida medicare patients in 2011

The algorithm written in R software [9] creates a survival object which adds an indicator function to the response that identifies whether it is censored in the model or not. For the survival function we begin with the Kaplan-Meier (KM) estimate procedure. This is a non-parametric maximum likelihood estimate (MLE) of the survival function based on the empirical probability density function of the proportion of patients who are not discharged on a particular day. That is, if we define d_j to be the number of patients who have failure time at x_j and n_j to be the number of patients at risk at time x_j (those who have already been discharged are no longer at risk), then we can define the estimate of the hazard function as

$$\hat{\lambda}_j = \frac{d_j}{n_j} \tag{8}$$

and an associated estimate of the survivor function is

$$\hat{F}(t) = \prod_{j|x_j<t} \left(\frac{n_j - d_j}{n_j}\right) = \prod_{j|x_j<t} \left(1 - \hat{\lambda}_j\right) \tag{9}$$

These two calculations are simple definitions of the population survival functions and population hazard functions and can be calculated easily. For the analysis of the effect of covariates on the response variable, categorical variables on the different factor levels are added using the same calculations.

However when adding an ordinal covariate, for example age, we need to consider a flexible modeling approach. We use CPF model on the length of stay data for different covariates. The CPH model is the most widely used model [10], and is described as an alternative to the AFT model [7]. The CPH model is a discrete failure time regression model which specifies a linear log odds model for the hazard probability at each potential failure time. This is done by creating a baseline hazard function $\lambda_0(t)$ and identifying a linear component $\boldsymbol{z\beta}$ with respect to the log of the odds provided by the hazard function to incorporate covariates $\boldsymbol{z}$ into the model. A sample of data with estimated hazards and estimated survival functions is given below.

Since a saturated CPH model shows that interactions on many variables are non-significant, we consider the reduced CPH model with some main effects and some corresponding significant interaction terms as given in the Fig. 2.

All of the variables included in the reduced model in Fig. 2 are the explanatory variables with at least one significant level of interaction. No interaction with sex is found to be significant hence we remove sex from the model. We also remove the three-way interaction between age, diagnosis, and respiratory charge, as well as the pairwise two-way interactions between the three variables since none are significant. The highest level of 4-way interaction is significant for this model. This relationship is important for future model building and for interpretation of the coefficients in the model.

Table 1. A sample of estimated hazard and estimated survival probability calculations from the data

Length of stay in days (t)	n_t	d_t	$\hat{\lambda}_t$	$\hat{F}(t)$
0	111,068	1,357	0.012218	1
1	109,711	7,661	0.069829	0.987782
2	102,050	15,775	0.154581	0.918806
⋮	⋮	⋮	⋮	⋮
15	6,726	908	0.134999	0.060557
16	5,818	704	0.121004	0.052382

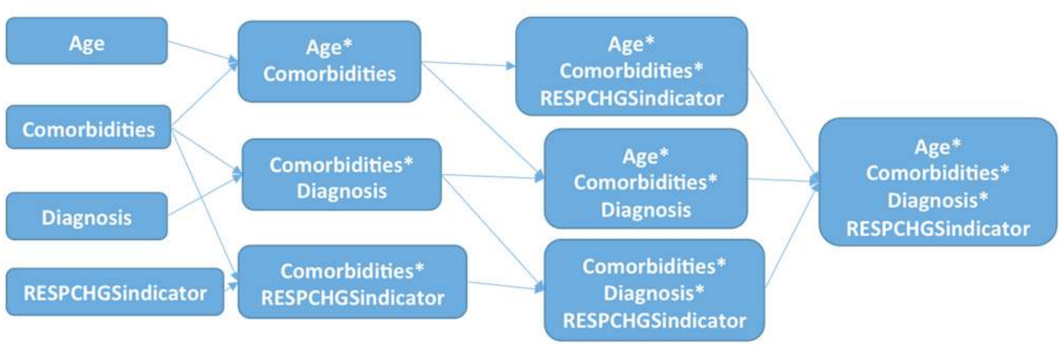

Fig. 2. CPH reduced model with significant two-way, three-way, and four-way interaction

Fig. 3 shows an example of the survival function using CPH model for a 65-year-old female with pneumonia but no respiratory charges at two different levels of comorbidities. Also, note that the same figure is valid with the survival function for a 65-year-old male with pneumonia, but no respiratory charges with two levels of comorbidities, since sex is not an influential factor.

As for the comparison of no comorbidities with major comorbidities when we compare the survival functions using CPH model for a 65-year-old with pneumonia and no respiratory charges, we see the survival functions are different. We see that 49.82% of these patients with no comorbidities will be discharged on or after the second day, meaning that half of the patients will have been discharged on or before the end of the first day. This is compared to 81% of patients with MCC being discharged after the second day. Patients with no comorbidities discharged on or after the seventh day are 0.75% of the population. This is compared to 23.44% of 65-year-old patients with pneumonia and no respiratory charges with MCC using CPH model are discharged on or after the seventh day. Notice that the survival curve levels out at 3% around 15 days for MCC. This is a projection on the patients with MCC, who will be in the hospital for a period of time longer than the 16 days, as we have right-censored data. When we increase the age of the patient to 85, the survival function shifts with all increased values. CPH model identifies age and comorbidity with interaction effect being present.

Now we consider an AFT model for identification of the interaction terms to be considered in the model. The AFT model begins with the same baseline as the CPH model, that is the same $\lambda_0(t)$, but instead of identifying a linear model $z\beta$ with respect to the log of the odds provided by the hazard function as CPH does, the AFT model identifies a model $z\beta$ that is proportional to T using lognormal link. This process uses the covariates as multiplicative factors. Since zeros are valid responses in our data we applied an additive correction factor of 1 for lognormal transformation. Fig. 4 describes the reduced AFT model for significant interaction factors.

For AFT model all explanatory variables are significant but not every interaction of covariates is. Respiratory charges are significant and present in every interaction effect. This indicates that if a patient went through a respiratory procedure it would significantly affect the influence of all other covariates. The saturated model with 5-way interaction is not significant for this model. For AFT model, the impact of comorbidity on length of stay is not influenced by sex. Since the AFT model includes all available variables it provides better estimates of the model parameters but is less robust than the CPH model. The CPH model behaves more accurately in extreme points on the survival functions.

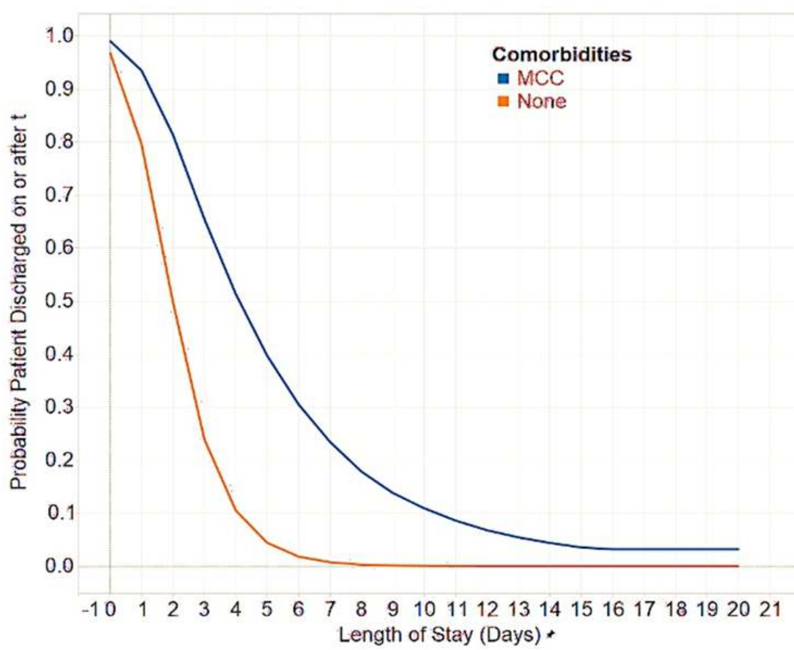

Fig. 3. Patients who are 65-years-old with PN, and no respiratory charges for both sexes using CPH model

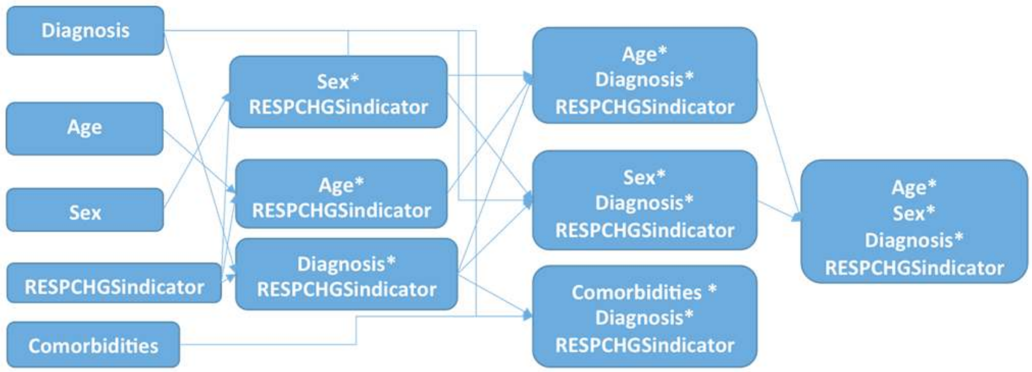

Fig. 4. AFT reduced model with significant two-way, three-way, and four-way interaction

Fig. 5 shows an example of the survival function using the AFT model for a 65-year-old female with pneumonia, at two different levels of comorbidities but no respiratory charges (on the left). Note the different curvature in the survival function using the AFT model for a 65-year-old male with pneumonia, with two different levels of comorbidities but no respiratory charges (on the right) since sex is a significant factor for this model. The curves for male patients drop more rapidly.

The criteria used for fitting the two models are given below in Table 2. In our model building we used 70% of the data to train the model parameters and used the remaining data as out of sample records to test how well the model acts on unobserved data. The sample is a simple random sample across all the variables.

The graphs in Figs. 6,7 and 8 are the out of sample data for the prediction using AFT and CPH models with KM estimates from the data censored after 16 days. We check for the goodness of fit [11] for three specific combinations of conditions for a patient based on a relatively healthy patient, a moderately ill patient, and a severely ill patient in out of sample data.

3.1 Relatively Healthy Patients

Fig. 6 is for estimated survival functions of two models for a relatively healthy group of patients. The KM line does not drop as quickly as the model functions do. This means in out of sample data the likelihood of a patient staying less than 3 days is much less than the models predict for relatively healthier patients. But, the models differ heavily after 3 days for this set of conditions. The relatively healthier patients are not well predicted by either of these two models.

3.2 Moderately Ill Patients

Fig. 7 is for the estimated survival functions of two models for a moderately ill group of patients with KM estimates. The KM line does not drop as quickly as the models do. This means in out of sample data the likelihood of a patient staying less than 3 days is still less than the models predict. However, this difference is not as severe as for relatively healthier patients. These set of conditions are more frequent in the out of sample data. However; the models behave very similarly for the given conditions.

Table 2. Diagnostic statistics for the fitted CPH and AFT models

CPH model		Significance	AFT model with lognormal link		Significance
Concordance	0.654	se = 0.001	Scale	2.09	NA
Rsquare	0.16	NA	Loglikelihood (model)	-209611.1	NA
Likelihood ratio test	13328	On 35 d.f. p=0	Loglikelihood (intercept only)	-215182.8	NA
Wald test	13907	On 35 d.f. p=0	Chisquare	11143.45	On 71 d.f. p=0
Score (Logrank test)	15017	On 35 d.f. p=0	Number of Newton-Raphson Iterations	4	NA

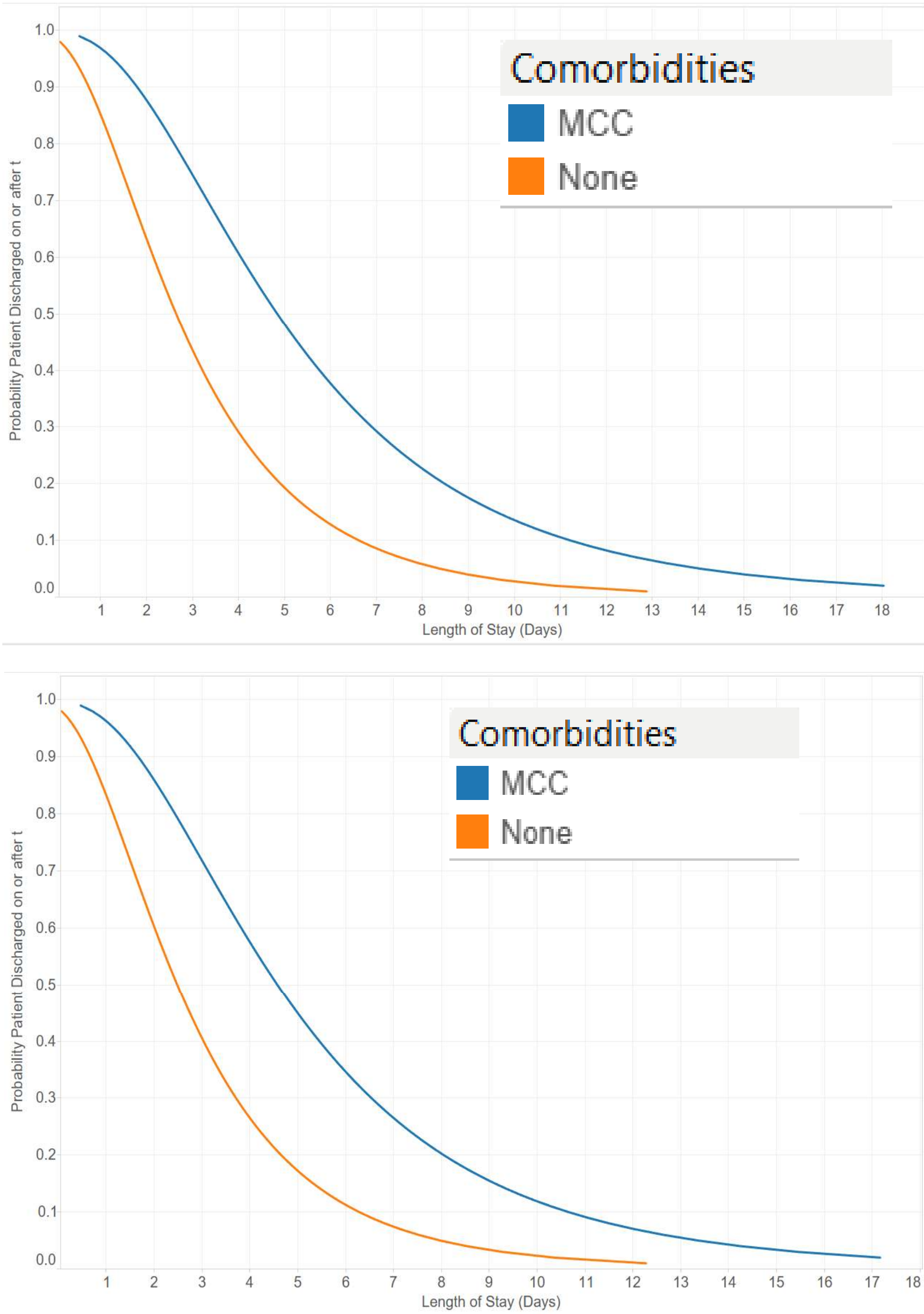

Fig. 5. Patients who are 65 years old with PN, and no respiratory charges with gender differences for female (top) and male (bottom) using AFT model

3.3 Severely Ill Patients

Fig. 8 is for estimated survival functions of two models for a severely ill group of patients with KM estimates. These three conditions are the worst scenario for this age group and represent a significant population in hospital staying for the Medicare group. The KM line does not drop as quickly as the models do but follows both models very closely. CPH performs better at the extreme values of the distribution of length of stay. AFT curve veers away from the KM line for longer lengths of stay.

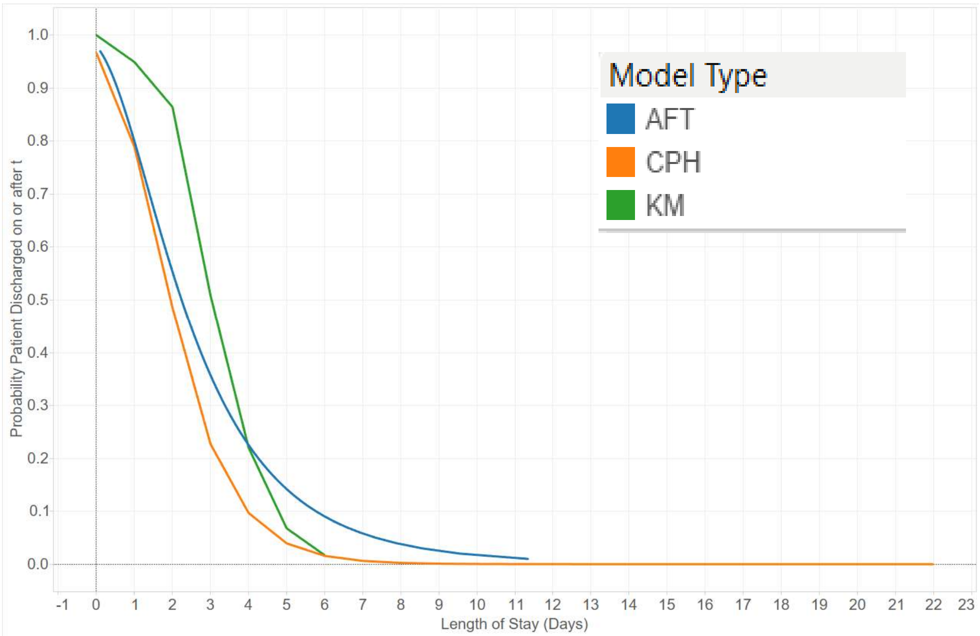

Fig. 6. Estimated AFT and CPH survival functions along with KM estimates of the observed out of sample values for a male patient who is 65-years old with AMI, no-comorbidity, and no respiratory charges

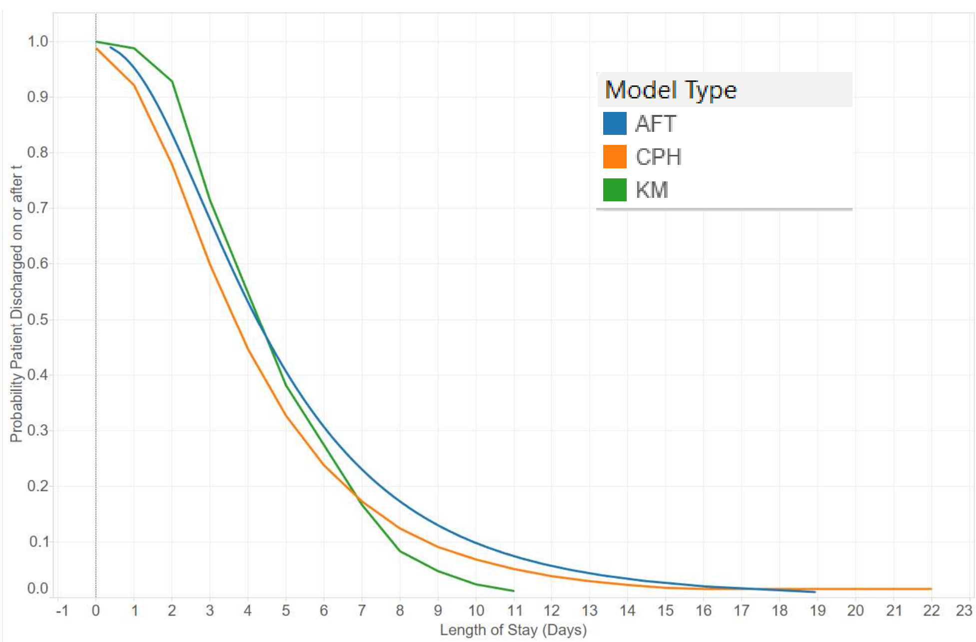

Fig. 7. Estimated AFT and CPH survival functions along with KM estimates of the observed out of sample values for a male patient who is 75 years old with HF, CC, and with respiratory charges

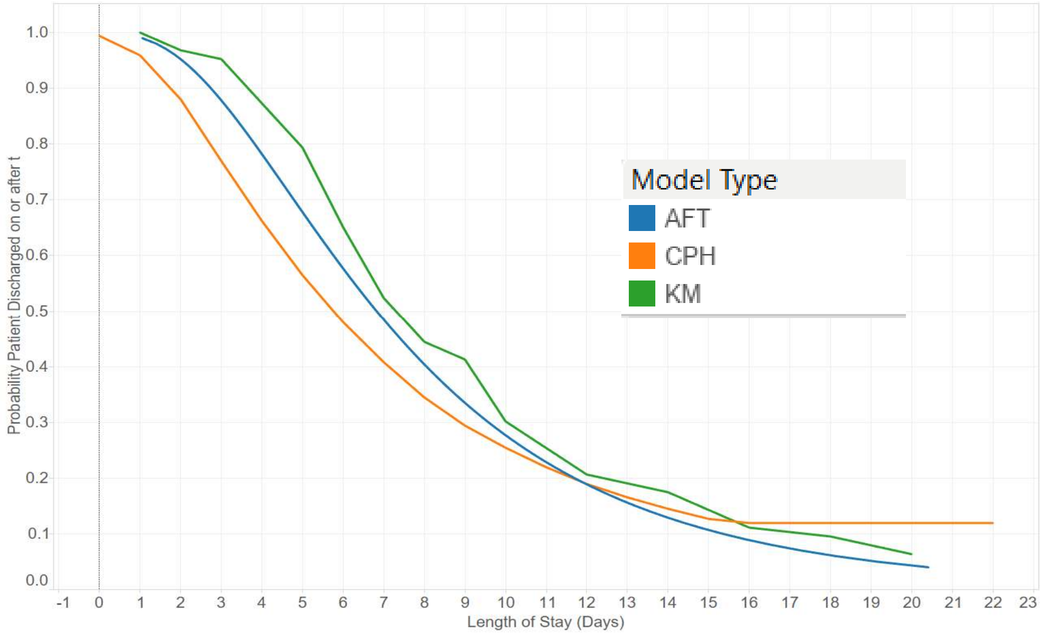

Fig. 8. Estimated AFT and CPH survival functions along with KM estimates of the observed out of sample values for a female patient who is 85 years old with PN, MCC, and with respiratory charges

The chi-square Table 3 shows that both models have a good fit since all the chi-square values are less than the critical value from a chi-square table for 14 degrees of freedom (d.f.).

Residual analyses done in Tableau [12] show the point estimates on responses using the AFT and the CPH models, which reveal the following heat maps in Fig. 9. If one is interested in estimating the prediction of the mean response for an individual, the AFT model predicts with more precision about the outcome than the CPH model. The CPH model generated a wider range of values and it worked better for the tail distributions for the survival function. The AFT model generated a tighter rage of values for the mean responses and it worked better for the middle on the survival function.

4. DISCUSSION

This analysis was completed using Microsoft Excel 2013 [13], R 3.2.0, and SQL Server 2012 [14]. The resulting implementation of the estimation of the coefficients of the models uses the *survival packages* in R. Censoring the data at 16 days is justified since the data has fewer observations for each passing day over 16. We use two powerful models, a parametric AFT model and a nonparametric CPH model. The available failure time data is rife with information, but is a challenge to use because the data has high variability at each combination level of the explanatory variables. KM estimates create the comparable data to be used by both models.

It is clear from Table 3 that both models fit our data extremely well since the goodness of fit tests concluded with non-significant chi-square values. The residuals are significantly small in most cells but some cells have higher errors due to the higher variability in combinations of diagnosis and comorbidity levels when respiratory charges are present. The heat maps in Fig. 9 indicate higher residual variables by darker shades.

Table 3. Goodness of fit test statistics for out of sample distributions for three ill patient types using AFT and CPH models

Chi-square goodness of fit statistics with the critical value of 23.6848, on 14 d.f.			
Patient	**65, Male, AMI, No Comorbidities, No respiratory charges (a)**	**75, Male, HF, CC, respiratory charges (b)**	**85, Female, MCC, respiratory charges (c)**
Model			
AFT	0.285604245	0.197916667	0.206388
CPH	0.10413797	0.113281953	0.290817

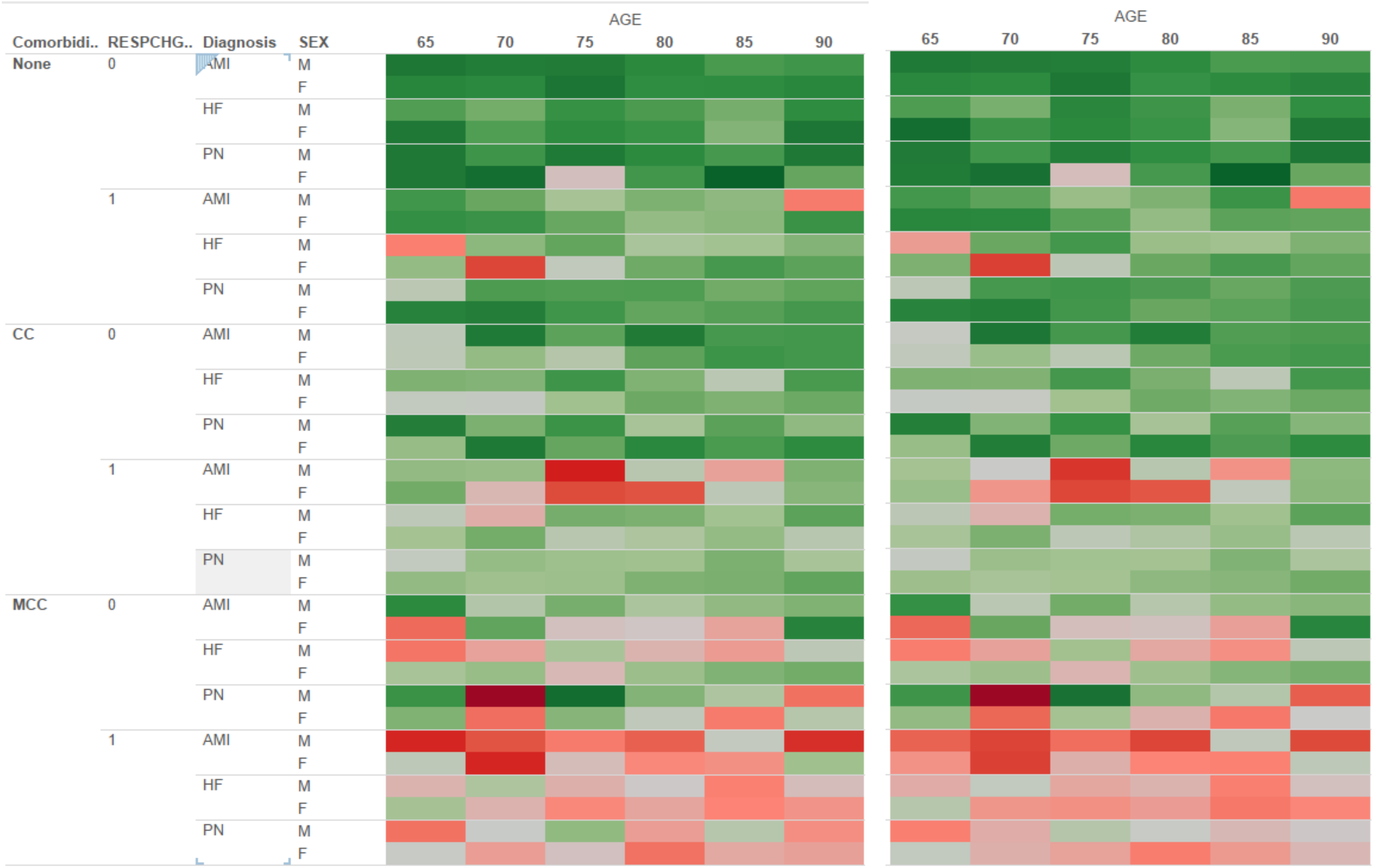

Fig. 9. Residuals for cross tabulation by sex, diagnosis, comorbidity level, and respiratory charges using AFT model (left) and CPH model (right)

A similarity between the CPH and the AFT models is that the predictions are very similar for the explanatory variable age. Older patients tend to stay longer independent of severity of illness. The CPH model does not identify sex as an important covariate. Other covariates such as diagnosis, comorbidities and respiratory charges are important for both models. When respiratory charges do not exist there is a sharp drop in the CPH model for similar patients with similar conditions. The respiratory charges play a significant role in predicting the length of stay. The effect is most extreme in the CPH model, where the absence of respiratory charges implies a sharp decrease in the longer length of stay. Both models do the predictions well, while the CPH model does a better job investigating the behavior of the survival function at the tail, but is less reliable for the center of the distribution of the function.

Both models tend to predict similar values when there is no comorbidity present and no respiratory charge is applied to the patients. As patient's complications increase the CPH model predicts better than the AFT model for extreme values. In general the AFT provides a smooth fitting of the KM estimates to the data. A complication in diagnosis of patients makes predictions more difficult for either of the models.

The first strength of this analysis is that when both models agree we are confident about our conclusion. When models disagree we look at the second strength. The second strength is that under certain circumstances one model is preferable than the other. We can apply each model for a restricted subset of variables and reliably trust the results under those conditions. The third strength of our research is that if sex is unknown at the time of the prediction, we can apply the CPH model which is not affected by the sex variable.

The first limitation of this research is that the disagreement causes confusion about influence of the variables. To give an example for HF patients with MCC condition the AFT will claim that increasing age increases the length of stay, whereas the CPH claims that increasing age decreases the length of stay. The second limitation is the difficulty in deciding which model to use for moderately ill patients because the CPH and the AFT work better on the relatively healthy patients or severely ill patients respectively. So for moderately ill patients we do not know which model to trust. The third limitation is that we are using billing data, which is collected after the patient is released from the hospital. Therefore the variable respiratory charges are not known at admission. These models cannot make prediction on the respiratory charges.

5. CONCLUSION

In recent years, a noticeable number of regulatory and oversight government agencies are using different statistical techniques to monitor payments to healthcare providers [15]. This research provides another method to explore how to screen the discharge of patients without a revisit within 30 days.

We have used two different models in order to predict length of stay data of patients according to specific data. It is necessary to appreciate the complexity of the data collection, human errors involving the judgment on a patient's discharge, and false positive or false negative detected by the medical tests. Despite having wide variability in the data we find models that describe observed values well. We see big disagreements between the two model types for some combinations of covariates. This generates the big question of which model predicts better than the other. This highlights the importance of using multiple model types when analyzing any data including failure time data, which we use here to model the length of stay in a Florida hospital for Medicare patients.

DISCLAIIMER

The title of this manuscript was presented in the conference "The Annual American Statistical conference". Available link is http://www.nlplogix.com/nlp-logix-stats-leaders-present-american-statistical-conference date feb18-2016 San Diego, California.

ACKNOWLEDGEMENTS

The authors would like to acknowledge three anonymous referees for their valuable comments to the readability of the paper.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Hospital Readmission Reduction Program (HRRP). Available:http://cms.gov/Medicare/Medicare-Fee-for-Service-Payment/AcuteInpatientPPS/FY-2012-IPPS-Final-Rule-Home-Page-Items/CMS1250103.html

2. CMS.gov: Access to CMS Data & Application; 2013.
Available:http://www.cms.gov/Research-Statistics-Data-and-Systems/CMS-Information-Technology/AccesstoDataApplication/index.html
3. AHCA: Florida Agency for Health Care Administration; 2014.
Available:http://www.ahcancal.org/Pages/Default.aspx
4. Webster B, Sen P. A prediction model for length of stay in florida hospitals: Big data analysis. Journal of Basic and Applied Research International. 2015;8(4):304–313. International knowledge Press.
5. Wei LJ. The accelerated failure time model: an useful alternative to the cox regression model in survival analysis. Statistics in Medicine. 1992;11:1871–1879.
6. Cox DR. Regression models and life tables. Journal of the Royal Statistical Society (Series B). 1972;34(2):187–220.
7. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. John Weily & Sons; 2002.
8. Le CT, Grambsch PM, Louis TA. Association between survival time and ordinal covariates. Biometrics. 1994;50:213–219. International Biometric Society.
9. Chiou SH, Kang S, Yan J. Fitting accelerated failure time models in routine survival analysis with R package aftgee. Journal of Statistical Software. 2014;61(11).
10. Therneau T. Survival: A Package for Survival Analysis in S. R package version 2.37-7.
Available:http://CRAN.R-project.org/package=survival
11. Novak P. Checking goodness of fit of the accelerated failure time model for survival data. WDS10 Proceeding of Contributed Papers, Part 1. 2010;189-194.
12. Tableau 8: The official guide, 1st Edition: McGraw-Hill Osborne Media; 2013.
13. Excel – Microsoft product 2013: Microsoft Office; 2013.
14. Lane W R, Looney SW, Wansley JW. An application of the cox proportional hazard model to bank failure. Journal of Banking & Finance. 1986;10(4):511–531.
15. Woodard B. Fighting healthcare fraud with statistics. Significance. 2015;12(3). Royal Statistical Society.