Cardiac Catheterization Outcomes and Their Association with Body Mass Index and Congestive Heart Failure Severity

Larry Chen, Justin Lim, Kevin Ordet, and Nagaprasad Rudrapatna

December 9, 2022

1 Introduction

1.1 Background Information

Cardiovascular diseases (CVD) are the leading cause of death worldwide and have been the leading cause of death in the United States since 1921 (Sawyer, 2021). The term “cardiovascular disease” applies to a wide range of heart-related ailments, including hypertension, coronary heart disease and cerebrovascular disease (i.e., strokes), and heart failure. While improvements in treatment methods have significantly lowered the number of deaths caused by cardiovascular diseases annually, understanding the key risk factors associated with cardiovascular diseases is crucial to the prevention of the disease.

Among these risk factors, cigarette smoking is a well-known predictor of CVD morbidity and mortality. A study led by Dr. Jane Ferrie, from the Department of Epidemiology and Public Health at University College London, studied the impact of common cardiorespiratory risk factors as predictors of mortality and concluded that smoking was the strongest predictor of mortality among women and a significant predictor of mortality for men (Ferrie et al., 2009). However, smoking itself is the cause of approximately 8 million premature deaths globally each year (GBD 2017 Risk Factor Collaborators, 2018), with the toxic and carcinogenic chemicals present in most tobacco products substantially increasing the risk of cancer and serious diseases (including cardiovascular diseases). The extent of the association between smoking and deaths due to cardiovascular diseases is therefore unclear.

Therefore, what could prove more useful would be understanding the association between a patient’s survival time and the combined effect of smoking with specific, well-known risk factors such as a patient’s body mass index (BMI) and the severity of their congestive heart failure (the inability of the heart to pump blood sufficiently to the rest of the body, such as during strenuous physical activity). This study acknowledges that, realistically, many patients likely suffer from a combination of key risk factors associated with cardiovascular disease, such as patients being overweight, smoking, and suffering from hypertension (high blood pressure). In our analysis, we hope to determine the extent to which patient outcomes (namely, death) are associated with a combination of risk factors.

1.2 Research Aims

Past studies have successfully explored the association between combinations of CVD risk factors and negative outcomes among patients in the past. For instance, a study conducted by Baghdadi et al. explores the relationship between risk factors among patients with another well-known CVD risk factor–rheumatoid arthritis–and concluded that factors, like hypertension, diabetes, and smoking, increased the risk of CVD in patients with rheumatoid arthritis (2015). This study uses survival analysis techniques to conduct a similar analysis with patients who smoke, in order to address the following research questions:

1. Is there an association between the severity of a (smoking) patient's congestive heart failure and survival time?
2. Is there an association between a (smoking) patient's body mass index and survival time?

1.3 Data Description and Key Variables

The original dataset for this report described adult patients who received cardiac catheterization procedures at the Duke University Hospital between 1985 and 2013. Cardiac catheterization is a medical procedure that involves inserting a thin tube into the patient's heart for either interventional or diagnostic purposes. This raw dataset included exactly one record per procedure, for a total of 83,320 procedures and 39,098 patients. Out of these patients, 21,492 are categorized as smokers, which is our target for our analysis.

1.3.1 Response Variable and the Right-censoring of Data

In this survival analysis, the response variable is the number of days from the catheterization to the last date that the patient was known to be alive. While patients with more than one catheterization had multiple records, we chose to only consider the first catheterization procedure to standardize the survival duration across all patients (see Methodology for further discussion). As the dataset only contains catheterization procedures occurring from 1985 to 2013, the data for patients who had a procedure around December 2014 is likely right-censored, as the duration of survival (time to last known alive status) is longer than the censoring time. This is demonstrated in Figure 1.

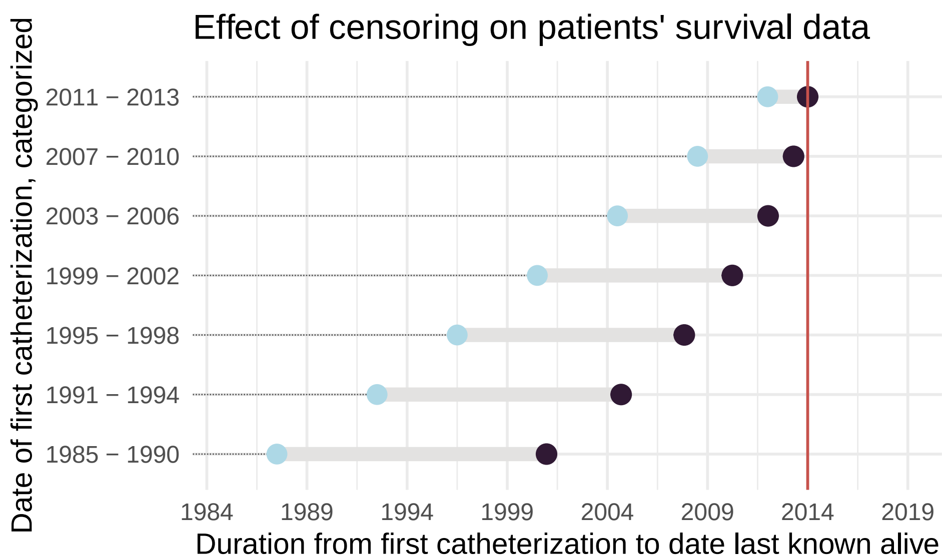

Figure 1: Red line shows the censoring time of December 2014

This visualization describes the effect of censoring which occurs in December 2014 on patients' survival data, one year after the date of the last observation included in the dataset. The plot shows patients categorized by the year of their first catheterization and uses the mean dates and duration values for each data point in the plot. Hence, it shows that the 2011-2013 group is on average affected by the right-censoring, while the 2003-2006 and 2007-2010 groups are likely to be affected by the right-censoring as well due to their proximity to the red line above. It is important to acknowledge that the censoring may be due to other factors as well (apart from the administrative censoring in 2014). For example, data for patients who no longer follow-up after a certain point (but before 2014) but who are still alive is also censored.

1.3.2 Key Predictor Variables

Table 1: Description of Key Covariates

Variable Name	Type	Description
Congestive Heart Failure Severity (CHFSEV)	Categorical	Describes degree of congestive heart failure (CHF) severity that a patient is suffering, ranked in ascending order from Class I to Class IV, with no congestive heart failure ('None') as the baseline
Body Mass Index (BMI)	Numeric	Numerical quantity obtained using the formula: 10000 * Weight (in kilograms) / Height^2 (in centimeters squared)
Body Mass Index Category (BMI_G)	Categorical	Classifying patients into the categories of underweight (< 18.5) , healthy weight (18.5-25; baseline category), overweight (25-30), and obese (> 30) based on the CDC guidelines

In this analysis, an emphasis is placed on the severity of a patient's congestive heart failure, as we are interested in this risk factor's relationship with smoking, which is also known to have an impact on physical health and endurance. According to a paper on congestive heart failure, heart failure itself can "severely decrease the functional capacity of a patient and increase mortality risk" (Malik et al., 2019). In this dataset, patients are categorized by how severely the patient's CHF condition limits their physical activity. We also chose to focus on the patient's body mass index (BMI), which is a well-known medical screening metric that uses the ratio of a patient's height to their weight to estimate if a patient has a healthy body fat percentage. In our analysis, we used the CDC guidelines on BMI (as the hospital is located in the US) to classify patients into the categories of underweight, healthy weight, overweight, and obese.

1.4 Hypotheses

We hypothesize that for patients with a history of smoking who underwent cardiac catheterization, those with normal BMI are expected to live longer than patients who are within the underweight, overweight, or obese BMI categories. We also hypothesize that for the same patients, those with less severe congestive heart failure are expected to live longer than those with a high severity for congestive heart failure, with survival time decreasing based on increased severity.

1.5 Exploratory Data Analysis

The Kaplan-Meier curves (Figures 2 and 3) describe the difference between the probability that a patient's last known alive date reaches up to a certain number of days, based on the different categories that they belong in. In the first plot, we observe that generally, as congestive heart failure severity worsens, the rate of decline in survival probability increases. This is mostly in line with our initial hypothesis (the curves for groups 1 and 2 intersect at approximately 5000 days). For the second plot, we observe that the underweight patients have the lowest survival probability, followed by patients with a healthy weight, and overweight and obese patients have the highest survival probability. This trend for the underweight category is in line with our hypothesis, but the trends for the overweight and obese categories seem to be different from our hypothesis.

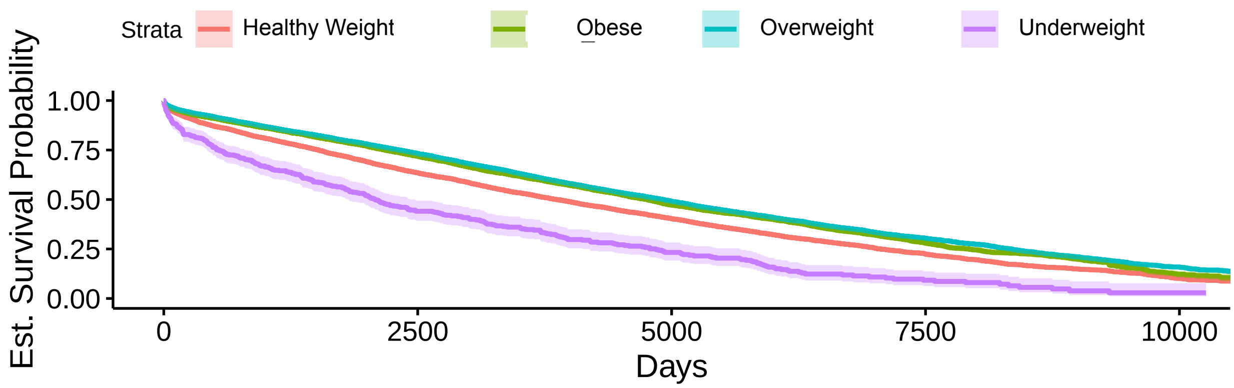

Figure 2: Survival Plot Based on BMI

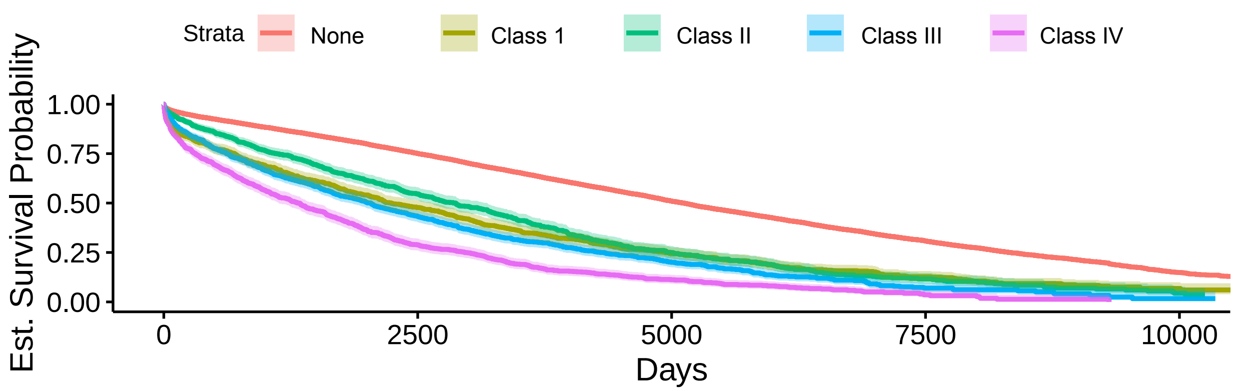

Figure 3: Survival Plot Based on CHF Severity

2 Methodology

We decided to restrict our analysis to patients who have a history of smoking. The motivation for this decision was derived from previous studies in the medical literature which demonstrated an association between smoking and heart disease. Furthermore, we decided to only account for the first cardiac catheterization procedure that each patient had. We made this decision to simplify the analysis and remove the hierarchical structure of the data (i.e., patients with multiple procedures). By only considering each patient's first procedure, we can treat the observations independently (i.e., independence assumption satisfied). Our approach to addressing the missingness (see Appendix A) in the dataset was to observe the patterns of the missing data and apply an approach suited to each variable. We observed that for the variable corresponding to the patient's race, for patients with multiple catheterizations, if the patient's race was not included in the first catheterization, it was omitted for all subsequent catheterizations. This suggests that the patient's race was a variable that was systematically omitted for specific patients and could not be classified as missing at random (e.g., the patient could have specified that they preferred race to be left out). Therefore, we chose to assign a new race category, corresponding to this systematic omission. As for the data missing from the congestive heart failure severity and coronary dominance, we did not observe a pattern within the missing data and chose to assume that the data were missing completely at random. While this assumption could lead to bias if the data was in fact not missing completely at random, we felt that the small proportion of missing data (approximately 1.4%) would not cause our analysis to be overly biased or lose statistical power. Therefore, we omitted the remaining observations with missing data for a complete case analysis.

After completing the data cleaning phase (which included recoding/modifying the covariates of interest and other controlling variables, namely the patient's age category and the year of his/her procedure; for example, we classified age into four broad groups–18-29, 30-49, 50-64, and >65–based on our intuition about when major health changes occur instead of sticking with the initial 13 groups to reduce the number of fitted coefficients in the final model), we created the final dataset. This final dataset includes 21,180 observations (patients with a history of smoking) and 15 columns (variables), namely the number of days since the catheterization to the last date that the patient was known to be alive, whether the patient died or was still alive at the end of follow-up (censored in 2014), the patient's age at the time of the catheterization in years, the patient's race, the patient's gender, the patient's BMI category, the patient's congestive heart failure severity, the number of catheterizations, the year of the catheterization, the patient's acute coronary syndrome status, the patient's history of cerebrovascular disease, the patient's history of diabetes, the patient's history of hypertension, the number of previous myocardial infarctions, and the patient's coronary dominance. We decided to fit an accelerated failure time (AFT) model, specifically with a log-logistic distribution, and selected the important covariates (the patient's congestive heart failure severity and BMI; see Key Predictor Variables for justifications) and additional effects to be controlled for based on our knowledge of clinical data (and the medical literature)–the key assumption being that patients likely suffer from a combination of several risk factors associated with cardiovascular disease. Using mathematical notation, we can express our final model as

$$\log(T_i) = \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \beta_3 x_{3i} + \beta_4 x_{4i} + \beta_5 x_{5i} + \beta_6 x_{6i} + \beta_7 x_{7i} + \beta_8 x_{8i} + \ldots + \beta_{30} x_{30i} + \epsilon_i$$

where i is the index for the patient (ranging from 1 to 21,180), T_i is the predicted survival time of ith patient with a history of smoking. In terms of the covariates, x_{1i} through x_{3i} indicate the ith patient's body mass index category, x_{4i} through x_{7i} indicate the ith patient's congestive heart failure severity, and x_{8i} through x_{30i} are the controlled patient demographic and medical factors.

We believe our modeling approach is reasonable for this analysis for multiple reasons; it is also worth mentioning that we used the `Surv()` function in R to construct our response variable so that the right-censoring was accounted for. Firstly, AFT models are appropriate for survival data, particularly when the proportional hazards assumption is violated (and hence the more popular Cox proportional hazards model is inappropriate). Additionally, since our research questions are inferential in nature and the target audience is clinicians (who do not have the strongest statistics background), we decided that an AFT model, with its simpler interpretations, would be better than a Cox proportional hazards model with several interaction effects (between explanatory variables and time to ensure that the proportional hazards assumption is satisfied). We considered Weibull, log-normal, and log-logistic distributions for the distribution of T_i and ultimately decided to use the log-logistic distribution after examining a Kaplan-Meier plot of the censored residuals against the survival functions of these distributions (see Appendix B) because it provided the best fit to the data. The plot in Appendix B illustrates that the log-logistic distribution is still unsatisfactory however as there is significant deviation from the expected curve in several regions. This means that the distributional assumption (that the assumed parametric form of the model adequately represents the data) of ϵ_i following a logistic distribution is violated. Linearity (in the coefficients) can also be assessed based on these Kaplan-Meier plots. The linearity assumption is likely violated because the Kaplan-Meier plot of the censored residuals includes strange patterns (i.e., unexpected behavior). We believe that these violations of the modeling assumptions may be the result of poor variable selection. The final AFT model may be missing important covariates (or transformations may be needed) which could account for patterns in the survival times. We did not conduct sensitivity analyses but believe they might be useful in further diagnosis of the problems. In the end, we decided to continue with this AFT model because we could not think of a better approach. Although several coefficients corresponding to controlling variables in the final model were not statistically significant (based on their respective p-values) assuming an α level of 0.05, we decided to keep them all since we believe that clinicians would be interested in the associations between survival time and congestive heart failure severity and body mass index, respectively, after accounting for the selected demographic and medical factors.

3 Results

The regression coefficients were estimated using the `survreg()` function in R. Table 2 displays the exponentiated coefficient estimates, as well as the corresponding 95% confidence intervals and p-values for the final model (see Appendix C for the full model output).

Table 2: Exponentiated Final Model Coefficients and 95% Confidence Intervals for Key Covariates

	Coefficient	Lower Bound	Upper Bound	P-Value
Intercept	2411.533	2198.967	2644.648	<0.01
Obese (BMI >30)	1.174	1.114	1.238	<0.01
Overweight (BMI 25-30)	1.273	1.215	1.334	<0.01
Underweight (BMI <18.5)	0.576	0.495	0.669	<0.01
Congestive heart failure severity class I	0.542	0.489	0.601	<0.01
Congestive heart failure severity class II	0.630	0.573	0.691	<0.01
Congestive heart failure severity class III	0.452	0.413	0.493	<0.01
Congestive heart failure severity class IV	0.309	0.280	0.341	<0.01
Log(scale)	0.793	0.781	0.804	<0.01

Holding the cardiac catheterization patient's BMI category constant and after accounting for all patient demographic and medical factors (that were included in the final model):

- A smoking patient with Class I congestive heart failure (CHF) is expected to live 0.542 times as long as a smoking patient without CHF. We are 95% confident that a smoking patient with Class I CHF is expected to live between 0.489 and 0.601 times as long as a smoking patient without CHF.
- A smoking patient with Class II CHF is expected to live 0.630 times as long as a smoking patient without CHF. We are 95% confident that a smoking patient with Class II CHF is expected to live between 0.573 and 0.691 times as long as a smoking patient without CHF.
- A smoking patient with Class III CHF is expected to live 0.452 times as long as a smoking patient without CHF. We are 95% confident that a smoking patient with Class III CHF is expected to live between 0.413 and 0.493 times as long as a smoking patient without CHF.
- A smoking patient with Class IV CHF is expected to live 0.309 times as long as a smoking patient without CHF. We are 95% confident that a smoking patient with Class IV CHF is expected to live between 0.280 and 0.341 times as long as a smoking patient without CHF.

We can conclude that, holding the patient's BMI category constant and after accounting for patient demographic and medical factors, smoking patients with Class IV CHF are predicted to have (statistically; $\alpha = 0.05$) significantly lower survival times than those with all other classes of CHF. Further, smoking patients without CHF are expected to survive significantly longer than those with any class of CHF. The distinction between Classes I-III is less clear; we can only conclude that smoking patients with Class II CHF are expected to live longer than those with Class III CHF.

Holding the cardiac catheterization patient's CHF severity status constant and after accounting for all patient demographic and medical factors (that were included in the final model):

- An obese (BMI >30) smoking patient is expected to live 1.174 times as long as a normal weight (BMI 18.5-25) smoking patient. We are 95% confident that an obese (BMI >30) smoking patient is expected to live between 1.114 and 1.238 times as long as a normal weight (BMI 18.5-25) smoking patient.

- An overweight (BMI 25-30) smoking patient is expected to live 1.273 times as long as a normal weight (BMI 18.5-25) smoking patient. We are 95% confident that an overweight (BMI 25-30) smoking patient is expected to live between 1.215 and 1.334 times as long as a normal weight (BMI 18.5-25) smoking patient.
- An underweight (BMI <18.5) smoking patient is expected to live 0.576 times as long as a normal weight (BMI 18.5-25) smoking patient. We are 95% confident that an underweight (BMI <18.5) smoking patient is expected to live between 0.495 and 0.669 times as long as a normal weight (BMI 18.5-25) smoking patient.

There is statistically significant ($\alpha = 0.05$) evidence that, holding the patient's CHF severity status constant and after accounting for all patient demographic and medical factors, obese and overweight patients with a history of smoking are predicted to survive longer than normal weight smoking patients. Conversely, we are able to conclude that underweight patients with a history of smoking are expected to have significantly shorter survival times than all others. Although the estimate for overweight smoking patients is larger than that for obese smoking patients, there is insufficient evidence to conclude that one group survives longer than the other.

4 Discussion

4.1 Conclusion

Diving into our model results, we can observe the relationships between a smoking patient's BMI and CHF severity, respectively, and his/her survival time after a cardiac catheterization. For BMI, we found that the survival time of underweight patients is expected to be shorter when compared to normal BMI patients, and that overweight patients are expected to have a longer survival time than normal weight patients. This is not quite in line with our initial hypothesis that for smoking patients who underwent cardiac catheterization, those with normal BMI have longer survival time than patients who are within the underweight, overweight, or obese BMI categories. This was our hypothesis because we believed that those with a normal healthy body mass index tend to be healthier individuals than those with low or high body mass index. This hypothesis was correct for underweight patients, however overweight and obese BMI patients have a longer survival time than normal BMI patients. So why is this the case? How are overweight and obese BMI smoking patients living longer than normal BMI smoking patients? We believe that gender plays a potential role.

In our sample, a small minority of BMI categories are female–roughly 27% for normal healthy weight and obese and approximately 19% for overweight. However, in the underweight BMI category, about 44% of patients are female (see Appendix D). We believe that there is a possibility that females could be more likely to be impacted by heart disease after cardiac catheterization compared to men, resulting in shorter survival time. Dr. Tariq of Lenox Hill Hospital had similar results from his research on the after-effects of cardiac catheterization. In Dr. Tariq's research, he found that not only are "underweight patients ... five times more likely to die than obese patients", but that "length of stay for underweight patients was more than double that of normal-weight patients" (2017). Even more so, Dr. Tariq stated that he also found that women were predominantly underweight and how that potentially breaks the bias of "coronary heart disease disproportionately affect[ing] men more than women" (2017).

Outside of gender, we believe that another potential reason overweight patients are expected to have longer survival times than healthy patients is that healthcare physicians are more likely to believe that overweight/obese patients are at higher risk and thus recommend more extreme treatments, whereas physicians may be less likely to view underweight or healthy patients as being at high risk of coronary heart disease. Zafrir et al. is another researcher who found overweight and obese patients to have longer survival time after cardiac catheterization but that "after dividing the population referred for cardiac catheterization into distinct indications ... the inverse relationship ... was noted in overweight and obese patients" (2018).

Aside from BMI category, another variable we investigated was severity of congestive heart failure on survival time for cardiac catheterization patients. Our results for this variable aligned with our hypothesis. Smoking patients with any class of CHF severity had a shorter survival time than those with no severity of congestive heart failure. We see that patients with congestive heart failure severity classes I and II are similar in terms of survival time, but patients with class III heart failure severity are expected to have shorter survival time than classes I and II, and patients with class IV severity are expected to have the shortest survival time. We believe that this makes sense in a medical sense. Individuals with symptoms of congestive heart failure that cause issues with physical activity are those that are likely to have a shorter survival time compared to those with no symptoms of congestive heart failure.

4.2 Limitations and Future Directions

Our modeling framework does have multiple limitations. To start, the log-logistic distribution was our best option, but it still did not fit our data well (meaning that the modeling assumptions were violated). This means that our results (inference) and subsequent conclusions are not necessarily reliable. Additionally, as mentioned in our methodology, our missing completely at random assumption may be inappropriate, and MICE (multiple imputation via chained equations) or other missing data approaches (which presuppose a different missingness mechanism) may be more appropriate than the complete case analysis we conducted in this analysis. In terms of our group of interest–smokers, we were only given data on whether the patient was a smoker. Thus, we had no information on how much the patient smoked, how many years they have been smoking, and the severity of their smoking habits. For our predictor variable of BMI group, we cut our groups into those under a BMI of 18.5 as underweight, between 18.5 and 25 as normal healthy weight, 25 to 30 as overweight, and any above 30 as obese. One limitation is that patients do not fall uniformly into these categories (i.e., data is imbalanced). While the healthy weight and obese categories are roughly balanced, there are fewer underweight patients (they account for less than 2% of smoking patients) and significantly more overweight smoking patients. A similar phenomenon occurs for congestive heart failure severity, where the vast majority of smoking patients do not have any class of congestive heart failure, but classes I, II, and IV have relatively equal proportions (class III is most common). Our BMI groups also do not split into specific categories of obesity. Those that have morbidly obese BMIs may have different results from those are barely on the obesity range for BMI. It is important to acknowledge the limited generalizability of our conclusions since we are only looking at patients with a history of smoking who had a cardiac catheterization procedure between 1985 and 2013 at Duke University Hospital (location-specific; we do not consider patients with a history of smoking who had catheterization procedures elsewhere). There are therefore potential issues with selection bias in our data, as our sample is not necessarily representative of the population of interest (smokers) nor randomly selected. Furthermore, the independence assumption for the AFT model was only satisfied because we excluded all catheterization procedures beyond the first one. This decision unfortunately limited the scope of our analysis and hence the generalizability of our conclusions.

In our future work, we would love to investigate other confounding variables and compare our existing results to non-smoking populations. Our analysis focuses on smokers, but it would be helpful for healthcare physicians to see how BMI groups and congestive heart failure severity impacts survival time differently for smokers and non-smokers. We would also like to further investigate why overweight and obese cardiac catheterization patients seem to have longer survival time than healthy BMI patients, on average. Our two hypotheses revolve around gender and potential physician bias when treating overweight BMI patients.

4.3 Summary

Our work identifies the relationship between patient BMI and congestive heart failure severity on survival time for smoking patients that have had cardiac catheterizations. Our results indicate that underweight smoking patients are expected to have a shorter survival time and that overweight/obese BMI smoking patients are expected to have a longer survival time, when compared to healthy smoking patients. Additionally, we found that smoking patients with congestive heart failure severity are expected to have a shorter survival time than smoking patients with no congestive heart failure. We hope that our analysis brings insight for healthcare professionals and physicians who treat patients with CVD or patients that have had a cardiac catheterization.

5 Appendices

5.1 Appendix A. Missingness Plot

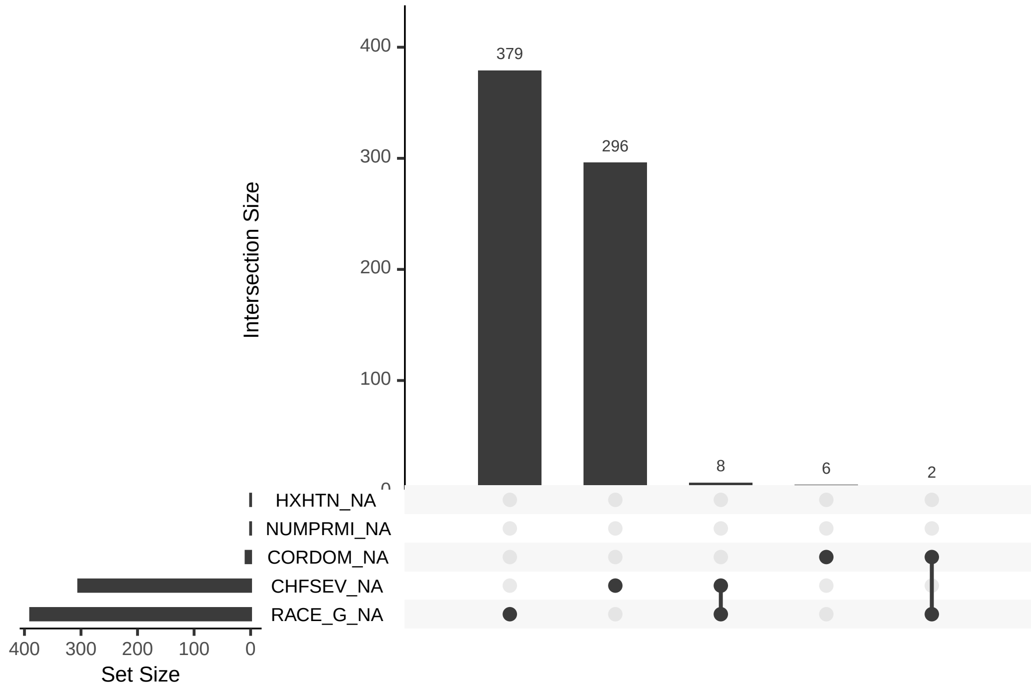

Figure 4: Diagnosing the Pattern in the Missingness

5.2 Appendix B. Model Assumptions and Diagnostics

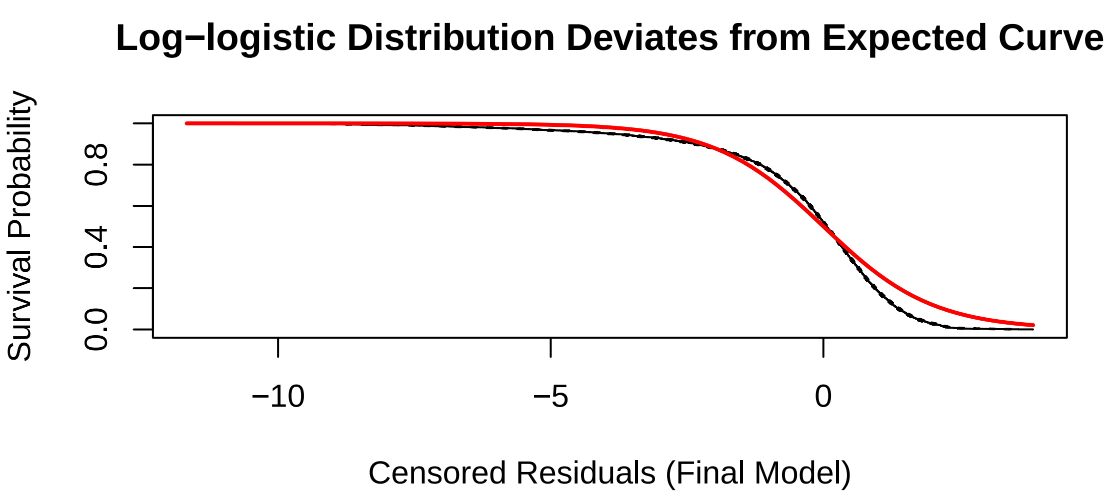

Figure 5: Kaplan-Meier Plot of Censored Residuals Against the Log-logistic Distribution

5.3 Appendix C. Supplementary Final Model Output

Table 3: Supplementary Exponentiated Model Coefficients and 95% Confidence Intervals

	Coefficient	Lower Bound	Upper Bound	P-Value
Age 18-29	3.095	2.045	4.685	<0.01
Age 30-49	2.977	2.803	3.161	<0.01
Age 50-64	1.937	1.854	2.024	<0.01
African-American	0.876	0.825	0.930	<0.01
Other race	0.954	0.857	1.062	0.3887
Missing race	0.860	0.733	1.008	0.0628
Female	1.027	0.980	1.075	0.2641
Total number of catheterizations	1.088	1.078	1.098	<0.01
Catheterization in 1991-1994	1.168	1.106	1.234	<0.01
Catheterization in 1995-1998	1.333	1.258	1.412	<0.01
Catheterization in 1999-2002	1.323	1.238	1.414	<0.01
Catheterization in 2003-2006	1.409	1.295	1.533	<0.01
Catheterization in 2007-2010	1.449	1.306	1.607	<0.01
Catheterization in 2011-2013	1.106	0.942	1.299	0.2189
ST-elevation myocardial infarction	1.042	0.985	1.102	0.1564
Non-ST-elevation myocardial infarction	0.986	0.908	1.071	0.7409
Unspecified myocardial infarction	0.957	0.865	1.058	0.3867
Unstable angina	1.036	0.985	1.090	0.1705
History of cerebrovascular disease	0.653	0.612	0.696	<0.01
History of diabetes	0.650	0.619	0.681	<0.01
History of hypertension	0.834	0.800	0.868	<0.01
Number of previous myocardial infarctions	0.832	0.806	0.859	<0.01
Right coronary dominance	1.054	0.981	1.132	0.1477
Balanced coronary dominance	0.985	0.869	1.116	0.812

5.4 Appendix D. Distribution of BMI for Females

Table 4: Analyzing BMI for Females

Gender	BMI_G	Percentage
Female	healthy weight	26.96
Female	obese	26.92
Female	overweight	19.00
Female	underweight	43.90

6 References

1) Sawyer, A., & Flagg, L. A. (2021). State Declines in Heart Disease Mortality in the United States, 2000–2019.

2) Ferrie, J. E., Singh-Manoux, A., Kivimäki, M., Mindell, J., Breeze, E., Smith, G. D., & Shipley, M. J. (2009). Cardiorespiratory risk factors as predictors of 40-year mortality in women and men. Heart, 95(15), 1250-1257.

3) GBD 2017 Risk Factor Collaborators. (2018). Global, regional, and national comparative risk assessment of 84 behavioural, environmental and occupational, and metabolic risks or clusters of risks for 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet (London, England), 392(10159), 1923.

4) Baghdadi, L. R., Woodman, R. J., Shanahan, E. M., & Mangoni, A. A. (2015). The impact of traditional cardiovascular risk factors on cardiovascular outcomes in patients with rheumatoid arthritis: a systematic review and meta-analysis. PloS one, 10(2), e0117952.

5) Malik, Ahmad, et al. "Congestive heart failure." StatPearls [Internet]. StatPearls Publishing, 2022.

6) Tariq, A. (2017, August 28). Underweight is associated with highest mortality and costs after cardiac catheterization. European Society of Cardiology Congress, 2017

7) Zafrir, B., Jaffe, R., Rubinshtein, R., Karkabi, B., Flugelman, M. Y., & Halon, D. A. (2018). Influence of body mass index on long-term survival after cardiac catheterization. The American Journal of Cardiology, 121(1), 113–119. https://doi.org/10.1016/j.amjcard.2017.09.028

7 Response to review

Introduction:

One issue I had with the introduction is occasional veering into causal language territory (or at least the implication that you will be conducting a causal analysis - for instance, from reading the sentence "the extent that smoking impacts deaths due to CV. . . " readers may assume that you will directly address this causal question when you do not; similarly, sentences such as "what could prove more useful would be understanding. . . the combined effect of smoking" give the same impression). The specific research aims are carefully worded to appropriately address only observational questions, but make sure that the same care is taken in the rest of the introduction.

We agree with your concern about casual language in the introduction. We made the following changes to the wording of the introduction: ". . . the extent of the association between smoking and deaths due to cardiovascular diseases is therefore unclear" "what could prove more useful would be understanding the association between a patient's survival time and the combined effect of smoking with specific, well-known risk factors such as a patient's body mass index (BMI) and the severity of their congestive heart failure"

The more significant problem I had with the introduction was that your research question is rather broad, but the dataset used are a specific subset of patients that might not be generalizable to your implied population of interest (in this case, smokers). The patients in the dataset are all at Duke for catheterization purposes, and smokers here might not be representative of smokers in general - what implications does this sort of conditional analysis have?

One point was returned after further review of the Limitations section. We also expanded on the implications of our simplified dataset: " It is important to acknowledge the limited generalizability of our conclusions since we are only looking at patients with a history of smoking who had a cardiac catheterization procedure between 1985 and 2013 at Duke University Hospital (location-specific; we do not consider patients with a history of smoking who had catheterization procedures elsewhere)."

Methodology:

However, I would have appreciated a bit more detail regarding the specific categorizations you used for some of your covariates (e.g., for age, I could only figure this out by going to the appendix. As well, since it wasn't of primary concern, why did you categorize it in the first place, and why did you choose the specific categories you did?).

We agree with your first comment and decided to include the following additional details about the variable categorizations:

“we used the CDC guidelines on BMI (as the hospital is located in the US) to classify patients”

“which included recoding/modifying the covariates of interest and other controlling variables, namely the patient’s age category and the year of his/her procedure; for example, we classified age into four broad groups–18-29, 30-49, 50-64, and >65–based on our intuition about when major health changes occur instead of sticking with the initial 13 groups to reduce the number of fitted coefficients in the final model”

One quick note is that there might also be censoring due to reasons besides the administrative censoring in 2014, but the manuscript seems to read as if the administrative censoring is the only source.

We acknowledged other reasons for censoring in the Response Variable and Right-censoring of Data section. For example, data for patients who no longer follow-up after a certain point (but before 2014) but who are still alive is also censored.

There is also a mistake in the model formulation as written - the use of the \hat{} implies that you’re providing a predicted survival time (this is fine). However, in this case it wouldn’t have an \epsilon_i term, since it is not observed and does not go into the prediction. In this case, removing the \hat{} on the T_i and \beta terms would correct the mistake.

We agree with this comment and removed the \hat{} on T_i and the \beta{} terms.

As a minor note, I’m not sure from the appendix plot in 5.2 that the log-logistic distribution “adequately represents” the data; there seems to be quite some deviation from the expected curve.

We agree that the phrasing “adequately represents” is too strong and decided to remove it. We instead stated that the log-logistic distribution provided the best fit to the data (compared to the Weibull and log-normal distributions). The plot in Section 5.2 illustrates that the log-logistic distribution is still unsatisfactory however as there is significant deviation from the expected curve in several regions. We therefore acknowledged that the assumptions for the AFT model were not satisfied.

Results:

Consider changing “alpha” to the symbol \alpha or describing it in terms of a priori significance level (i.e., in words) (I didn’t take off points for this; also consider specifying “statistically significant” instead of only “significant”).

We changed alpha to the symbol \alpha and also specified being “statistically” significant.

There does appear to be one misinterpretation - the manuscript says “obese and underweight patients ... are predicted to survive longer,” but this does not align with what is presented in the table (underweight lower expected survival; overweight higher expected survival compared to normal weight).

Thanks for pointing this out! We fixed this error. It should have stated “obese and overweight patients”.

Discussion:

There may additionally be issues here with selection bias, as we have a non-representative, non-randomly selected sample as the analysis dataset (you briefly touch on this topic, but do not go into much detail).

One point was returned after further review of the Limitations section. We modified our discussion to the following: “It is important to acknowledge the limited generalizability of our conclusions since we are only looking at patients with a history of smoking who had a cardiac catheterization procedure between 1985 and 2013 at Duke University Hospital (location-specific; we do not consider patients with a history of smoking who had catheterization procedures elsewhere). There are therefore potential issues with selection bias in our final dataset, as our sample is not necessarily representative of the population of interest (smokers) nor randomly selected.”