Mortality Risk Among Hemodialysis Patients with Low Serum Albumin and Cardiovascular Disease

Esteban Escobar[a]
[a]The University of California, Irvine, Department of Statistics

Abstract. Data from $N = 1,979$ hemodialysis patients were recruited for an observational study. The study aims to quantify the association between serum albumin and total mortality among hemodialysis patients. Furthermore, we want to assess if the association between albumin and mortality risk differs between patients with and without cardiovascular disease. Adjusting for all other covariates, we estimate the relative risk of mortality among hemodialysis patients is 57.3% higher comparing two sub-population, one of which are people with low albumin levels and the other are people with normal to high albumin levels [95% CI: (1.261,1.963), P-value < 0.0001]. We see similar results when we treat albumin as a time-varying covariate [estimate: 0.306, 95% CI: (0.246, 0.380), P-value < 0.0001]. Finally, using a partial likelihood ratio test, we have strong evidence against the hypothesis that there are no differences in the association between albumin and mortality risk depending on patients' cardiovascular disease status [p-value: 0.013]. That is, adjusting for all other covariates, we estimate the relative risk of death among patients with a history of cardiovascular disease is 26.3% lower comparing two sub-population differing in one unit change in serum albumin levels [95% CI: (0.568, 0.956) P-value: 0.0215]. We see similar results when treating albumin as a time-varying covariate [estimate: 0.371, 95% CI: (0.288, 0.477), p-value < 0.0001].

Keywords: Serum Albumin, Cardiovascular Disease, Cox Proportional Hazard Model, Time Varying Covariates .

1 Introduction

End-stage renal disease (ESRD) is a medical condition in which a person's kidneys cannot filter the toxic from one's body to the point where life can no longer adequately be sustained. Kidneys can filter wastes and excess fluids in the bloodstream, which are then excreted from urination, hence losing their filtration abilities, leading to dangerous levels of fluid, electrolytes, and wastes in the body. ESRD affects more than 1500 people per million in countries with a high prevalence, such as Japan, Taiwan, and the US [1]. According to data from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), over 850,000 persons in the United States are being treated for ESRD, and many more suffer from early-stage chronic kidney disease.

Serum albumin is a protein biomarker and a good measure of a person's nutritional status. It has been hypothesized that protein energy malnutrition (PEM) may be a marker of total mortality among hemodialysis patients. Hence it will be beneficiary if we can find an association between low albumin levels and the mortality risk among hemodialysis patients. Low albumin levels indicate malnutrition due to inadequate dialyzing, which can increase the risk of mortality among hemodialysis patients. If we can find an association, this will help nephrologists with an additional measure to monitor hemodialysis patients and potentially decrease the risk of mortality in these patients. Furthermore, it has also been hypothesized that an association between albumin and mortality risk may differ between patients with and without cardiovascular disease. If we can see this association, this is another measurement that nephrologists can look at to help decrease mortality among hemodialysis patients. This paper will attempt to answer the following scientific question of interest. Is there an association between serum albumin levels and the mortality risk among hemodialysis patients? Furthermore, we want to see if the association between serum albumin and mortality risk is different depending if a person has a history of cardiovascular disease. Finally, we will conduct a nested case-control design with four controls per case which will be conducted in Appendix C.

2 Statistical Methods

2.1 Data

To help address our hypothesis $N = 1979$, hemodialysis patients were randomly recruited to an observational study. Patients undergoing dialysis during December 2014 were eligible for the study and were recruited from dialysis clinics across the United States. Thinking about our sample scheme, we are sampling from people in the United States. Hence, we are not considering patients from outside the United States. Since hemodialysis is such an expensive process, everyone is covered. We are sampling from people diagnosed with End Stage renal disease (ESRD) or a similar disease requiring them to undergo hemodialysis therapy. Hence we are not sampling from healthy individuals. Overall, we believe we have a good sample scheme for our population.

Their serum albumin and multiple demographic and laboratory measurements were measured during recruitment. In addition, given the study's focus on albumin, we also measured and recorded their serum albumin levels every one to two months for up to one year following the start of the study. Other measurements and demographics are their age, sex, race group, smoking status, CVD history, diabetes status, total ESRD time, BMI, and an indicator of whether the patients appeared undernourished to the study nurse. We want to identify potential confounders and precision variables. Confounders are variables associated with our covariates of interest and our response. Potential confounders are age, sex, smoking group, BMI, history of CVD, and the total time of ESRD. Literature has shown an association between low serum albumin levels and these variables [2]. We can also see that these variables are tied to mortality rate; for example, as we age, the risk of mortality increases, or a current smoker will have a higher mortality risk than a person who never smoked. A precision variable is a covariate related to the response independent of the covariate of interest. We see that an indicator of whether the patients appeared undernourished to the study nurse could be a potential precision variable. If a nurse sees a person who is malnutrition, that can be a good indicator of an increase in mortality. Some other potential precision variables are race and diabetes status. African Americans have the highest death rates of any of America's racial and ethnic groups. This can be largely attributed to economic status, education, and occupation inequalities, all related to mortality risk [4]. Similarly, people with type 2 diabetes have a higher mortality risk.

2.2 Statistical Modeling

To help answer our scientific questions of interest, as discussed in Section 1, we apply suitable statistical modeling methods. Specifically, we will use a cox proportional hazard model where death is our event.

2.2.1 Mortality Risk and Albumin levels

In our first question, we desired to estimate the relative mortality risk among hemodialysis patients, comparing two sub-populations differing in one unit change in serum albumin and adjusting for all other covariates. We want to utilize Equation 1, which includes all the confounders and precision variables as discussed in Section 2.1. To determine the association between serum albumin levels and the mortality risk among hemodialysis patients, we can look at the coefficient associated with albumin levels and see if it is significant.

Similarly, we have time-varying covariates using the longitudinal data where we have serum albumin levels as a function of time. We can utilize the whole data using the same Equation but treating Albumin as a time-varying covariate. The interpretation, in this case, will change. Adjusting for all other covariates, a randomly sampled patient with serum albumin levels at any given time t has a different mortality risk than a similar patient at the same time t. We can determine the association similar to the baseline case, where we look at the p-value for significance.

$$\begin{aligned}\lambda(t) = \lambda_0(t)exp(\beta_1 albumin &+ \beta_2 \mathbb{1}(CVD = Yes) + \beta_3 Age + \beta_4 \mathbb{1}(Sex = Female) \\ &+ \beta_5 \mathbb{1}(Smoking = Former) + \beta_6 \mathbb{1}(Smoke = Current) \\ &+ \beta_7 BMI + \beta_8 ESRDtime + \beta_9 \mathbb{1}(undernourished = yes) \\ &+ \beta_{10} \mathbb{1}(Race = African\ American) + \beta_{11} \mathbb{1}(Race = Other) \\ &+ \beta_{12} \mathbb{1}(Diabetes = yes))\end{aligned} \tag{1}$$

2.2.2 Differences in Mortality Risk Among History of CVD

Recall the second question. We want to determine if the association between serum albumin levels and mortality risk among hemodialysis patients is different depending if a person has a history of cardiovascular disease. In this case, we want to utilize Equation 2 where we have an interaction between history of Cardiovascular disease and Albumin levels. To determine this association, we will conduct a partial likelihood ratio test to test if $\beta_{13} = 0$. Similarly, we can use the same method utilizing the whole dataset where albumin is a time vary covariate. We will use a partial likelihood ratio test to see if we have an association.

$$\begin{aligned}\lambda(t) = \lambda_0(t)exp(\beta_1 albumin &+ \beta_2 \mathbb{1}(CVD = Yes) + \beta_3 Age + \beta_4 \mathbb{1}(Sex = Female) \\ &+ \beta_5 \mathbb{1}(Smoking = Former) + \beta_6 \mathbb{1}(Smoke = Current) \\ &+ \beta_7 BMI + \beta_8 ESRDtime + \beta_9 \mathbb{1}(undernourished = yes) \\ &+ \beta_{10} \mathbb{1}(Race = African\ American) + \beta_{11} \mathbb{1}(Race = Other) \\ &+ \beta_{12} \mathbb{1}(Diabetes = yes) + \beta_{13} albumin \times \mathbb{1}(CVD = Yes))\end{aligned} \tag{2}$$

3 Results

3.1 Descriptive Statistics

To begin, we provided some descriptive statistics to help describe the basic characteristic of our sample used to address our questions of interest. Figure 1 compares the survival function to people with low albumin, people with a history of CVD, and a mixture of both. Looking at the lower left plot of Figure 1, we can see that people with low serum albumin and a history of CVD have a lower survival curve than the others. Other survival curves can be shown in Appendix A. Table 1 summarizes our data stratified by low serum albumin levels. Here if a patient has a serum albumin of less than 3.6 g/dL then they will be marked as someone with low Albumin [3]. Looking at the table, we see that not many people have low albumin levels, although the exact proportions seem different based on the history of CVD. We can also see that most patients are on the older side, where the median age is 63 years old. Similarly, summary tables are in Appendix A, where we stratified by CVD history.

Low Serum Albumin	No	Yes	Overall
	(N=1374)	(N=605)	(N=1979)
History of CVD			
No	554 (40.3%)	184 (30.4%)	738 (37.3%)
Yes	708 (51.5%)	372 (61.5%)	1080 (54.6%)
Missing	112 (8.2%)	49 (8.1%)	161 (8.1%)
Age			
Mean (SD)	59.3 (15.7)	62.6 (14.4)	60.3 (15.3)
Median [Min, Max]	62.0 [18.0, 93.0]	64.0 [17.0, 90.0]	63.0 [17.0, 93.0]
Sex			
Male	731 (53.2%)	282 (46.6%)	1013 (51.2%)
Female	643 (46.8%)	323 (53.4%)	966 (48.8%)
Smoking Group			
Never	700 (50.9%)	299 (49.4%)	999 (50.5%)
Former	376 (27.4%)	151 (25.0%)	527 (26.6%)
Current	187 (13.6%)	89 (14.7%)	276 (13.9%)
Missing	111 (8.1%)	66 (10.9%)	177 (8.9%)
BMI			
Mean (SD)	24.8 (5.74)	24.5 (5.80)	24.7 (5.76)
Median [Min, Max]	23.8 [13.4, 71.7]	23.3 [12.1, 54.9]	23.7 [12.1, 71.7]
Missing	25 (1.8%)	16 (2.6%)	41 (2.1%)
Total ESRD time			
Mean (SD)	3.28 (3.64)	2.30 (3.13)	2.98 (3.52)
Median [Min, Max]	2.06 [0, 20.2]	1.10 [0, 18.0]	1.80 [0, 20.2]
Undernourished			
No	1136 (82.7%)	379 (62.6%)	1515 (76.6%)
Yes	156 (11.4%)	188 (31.1%)	344 (17.4%)
Missing	82 (6.0%)	38 (6.3%)	120 (6.1%)
Race Group			
Caucasian	678 (49.3%)	355 (58.7%)	1033 (52.2%)
African American	565 (41.1%)	201 (33.2%)	766 (38.7%)
Other	108 (7.9%)	42 (6.9%)	150 (7.6%)
Missing	23 (1.7%)	7 (1.2%)	30 (1.5%)
Diabetes			
No	911 (66.3%)	349 (57.7%)	1260 (63.7%)
Yes	455 (33.1%)	255 (42.1%)	710 (35.9%)
Missing	8 (0.6%)	1 (0.2%)	9 (0.5%)

Table 1: Descriptive statistics of the patient's characteristics in the sample, stratified by low serum albumin. For discrete variables, we provide the count and percentage. We provide the mean, standard deviation, minimum, median, and max values for continuous variables.

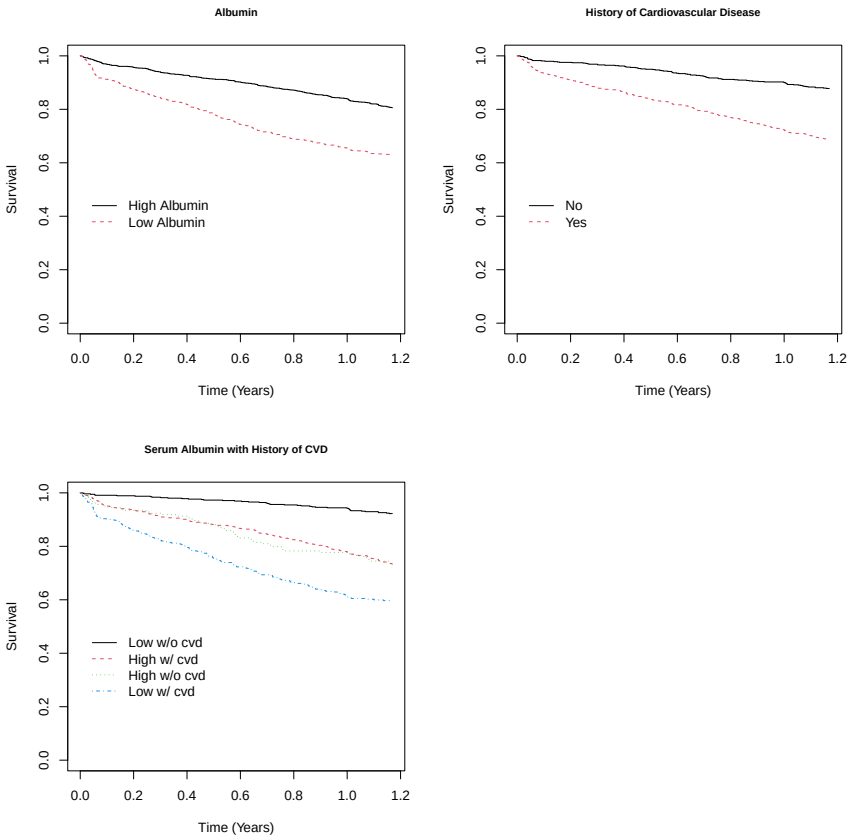

Fig 1: Plot of the survival function to total mortality among hemodialysis patients with low serum albumin, history of CVD, and a combination of both.

3.2 Mortality Risk and Albumin levels

The adjusted model was fit as defined in Section 3.2, we can see our estimates and corresponding 95% CI, Standard Error, and p-value as shown in Table 2. Looking at the baseline model, we see a p-value of less than 0.0001, and hence there is an association between serum albumin and mortality risk. We estimate the relative risk of mortality among hemodialysis patients comparing two sub-populaiton with similar cardiovascular history, age, sex, smoking status, BMI, total ESRD time, race group, diabetes status, and nourished status is 36.9% lower differing in one unite change in serum albumin levels. A better interpretation is to use the indicator of low albumin levels. We can say that adjusting for all other covariates, we estimate the relative risk of mortality among hemodialysis patients is 57.3% higher comparing two sub-population one of which are people with low albumin levels and the other are people with normal to high albumin levels.

Similarly, looking at Table 2 on the right columns, we have our estimated relative risk utilizing our full data and treating serum albumin as a time-varying covariate. We see that our estimates for serum albumin have a p-value less than 0.0001. Hence there is an association between the risk of mortality and low levels of serum albumin. Adjusting for all other covariates, randomly sampled patients with serum albumin at any given time t have about a 69.4% lower risk of death than a similar patient at the same time t with different albumin levels. Comparing these two models shows that the model with the repeated measurement has a smaller standard error. The reason might be that we treat albumin as a time function. Overall we still estimate an increase in the risk of death with low albumin levels.

Covariate	Baseline			Repeated Measurement		
	RR (95% CI)	SE	P-value	RR (95% CI)	SE	P-value
Serum Albumin (g/dL)	0.631 (0.503, 0.792)	0.116	<.0001	0.306 (0.246, 0.380)	0.111	<.0001
< 3.6 g/dL	1.573 (1.261, 1.963)	0.113	<.0001			
Cardiovascular Hist.						
No	Reference			Reference		
Yes	2.092 (1.599, 2.738)	0.137	<.0001	2.022 (1.547, 2.643)	0.137	<.0001
Age (10 years)	1.413 (1.291, 1.546)	0.046	<.0001	1.400 (1.281, 1.531)	0.045	<.0001
Sex						
Male	Reference			Reference		
Female	0.905 (0.725, 1.129)	0.113	0.3762	0.874 (0.700, 1.093)	0.114	0.2382
Smoking Status						
Never	Reference			Reference		
Former	0.944 (0.735, 1.213)	0.128	0.6528	0.956 (0.744, 1.230)	0.128	0.7271
Current	1.533 (1.130, 2.080)	0.156	0.0061	1.522 (1.120, 2.068)	0.156	0.0072
BMI (5kg/m^2)	0.894 (0.797, 1.002)	0.058	0.0548	0.885 (0.790, 0.993)	0.058	0.0367
ESRD Total Time	1.010 (0.976, 1.045)	0.017	0.5625	1.010 (0.976, 1.045)	0.017	0.5783
Undernourished						
No	Reference			Reference		
Yes	1.886 (1.482, 2.401)	0.123	<.0001	1.664 (1.316, 2.105)	0.120	<.0001
Race Group						
Caucasian	Reference			Reference		
African American	0.811 (0.642, 1.025)	0.119	0.0798	0.818 (0.648, 1.034)	0.119	0.093
Other	0.766 (0.481, 1.220)	0.237	0.2615	0.689 (0.433, 1.097)	0.237	0.1165
Diabetes Status						
No	Reference			Reference		
Yes	1.246 (0.991, 1.567)	0.117	0.0594	1.176 (0.934, 1.482)	0.118	0.1683

Table 2: The estimated relative risk of mortality among hemodialysis patients. Along with its corresponding 95% confident interval, standard error, and p values. The first column represents our estimates at baseline, while the second represents our estimates with time-varying covariates.

3.3 *Differences in Mortality Risk Among History of CVD*

Next, to determine if the association between albumin levels and mortality risk differs depending on CVD history, we will fit the model discussed in Section 2.2.2. Table 3 shows the estimated relative risk of death among hemodialysis patients along with their corresponding 95% confidence interval, standard error, and p-value. Recall that we fit an interaction between albumin and history of CVD, and using a partial likelihood ratio test to test our interaction, we have a p-value of 0.01304. Adjusted for all other covariates, we estimate the relative risk of death among patients with history of CVD is 26.3% lower comparing two sub-population differing in one unit change in serum albumin levels.

Similarly, utilizing the full dataset and treating serum albumin as a time-varying covariate, we can output their estimates, 95% confidence interval, standard error, and p-value as shown in Table 3 right column. Conducting a partial likelihood ratio test to test the interaction, we have a p-value of 0.0009. We see that in both the baseline and repeated measurement, the association between albumin and mortality risk differs depending on history of CVD. If we compare these two models, notice that the standard error for the coefficient of our interaction is smaller compared to the baseline model and the repeated measurement.

Covariate	Baseline			Repeated Measurement		
	RR (95% CI)	SE	P-value	RR (95% CI)	SE	P-value
Serum Albumin (g/dL)	0.404 (0.271, 0.604)	0.205	<.0001	0.163 (0.108, 0.246)	0.210	<.0001
Cardiovascular Hist.						
No	Reference			Reference		
Yes	0.246 (0.047, 1.286)	0.843	0.0966	0.117 (0.023, 0.598)	0.834	0.0100
Age (10 years)	1.412 (1.290, 1.545)	0.046	<.0001	1.398 (1.279, 1.529)	0.046	<.0001
Sex						
Male	Reference			Reference		
Female	0.913 (0.731, 1.140)	0.113	0.422	0.877 (0.701, 1.096)	0.114	0.2485
Smoking Status						
Never	Reference			Reference		
Former	0.941 (0.732, 1.210)	0.128	0.6358	0.955 (0.742, 1.228)	0.129	0.7184
Current	1.527 (1.125, 2.072)	0.156	0.0066	1.520 (1.119, 2.066)	0.156	0.0074
BMI (5kg/m^2)	0.891 (0.796, 0.998)	0.058	0.047	0.881 (0.787, 0.986)	0.057	0.0271
ESRD Total Time	1.014 (0.979, 1.049)	0.018	0.4436	1.015 (0.980, 1.050)	0.018	0.4030
Undernourished						
No	Reference			Reference		
Yes	1.888 (1.483, 2.402)	0.123	<.0001	1.645 (1.301, 2.080)	0.120	<.0001
Race Group						
Caucasian	Reference			Reference		
African American	0.815 (0.645, 1.030)	0.120	0.0871	0.828 (0.655, 1.046)	0.119	0.1131
Other	0.779 (0.490, 1.240)	0.237	0.2924	0.715 (0.449, 1.138)	0.237	0.1573
Diabetes Status						
No	Reference			Reference		
Yes	1.248 (0.993, 1.568)	0.117	0.0578	1.184 (0.940, 1.491)	0.118	0.1508
Serum Albumin (g/dL)						
- Cardiovascular Hist.						
No	0.404 (0.271, 0.604)	0.205	<.0001	0.163 (0.108, 0.246)	0.210	<.0001
Yes	0.737 (0.568, 0.956)	0.133	0.0215	0.371 (0.288, 0.477)	0.129	< .0001

Table 3: The estimated relative risk of mortality among hemodialysis patients. Along with its corresponding 95% confident interval, standard error, and p values. The first column represents our estimates at baseline, while the second represents our estimates with time-varying covariates.

4 Discussion

We saw in Section 3.2 the estimated relative risk of death among low serum albumin levels adjusting for all other covariates. In Section 3.3, we have the estimated relative risk of death among patients with cardiovascular history differing in one unit change in serum albumin levels. Both of these results share similar stories. Simply put, we estimate that higher serum albumin levels lead to a lower risk of death among hemodialysis patients, adjusting for all other covariates. Recall that our analysis mainly focuses on quantifying the association between albumin and mortality, which we have done. These findings are important because they can potentially help Nephrologists indicate among hemodialysis patients who are at higher risk for death and can take action before it is too late. However, we must remember that our conclusions and estimates in this study should be interpreted with the sample frame, namely patients undergoing dialysis during December 2014. Some limitations in our analysis come from the lack of some potential confounders, such as medication that the patients may have been on before the study. Some future work to consider is looking at other different laboratory measurements not in our data to see if there is an association with mortality among hemodialysis patients.

References

[1] Abbasi, M.A., Chertow, G.M., Hall, Y.N., 2010. End-stage renal disease. BMJ clinical evidence 2010, 2002.

[2] Erstad, B.L., 2021. Serum Albumin Levels: Who Needs Them? Annals of Pharmacotherapy 55, 798–804. URL: http://journals.sagepub.com/doi/10.1177/1060028020959348, doi:10.1177/1060028020959348.

[3] Nelson, C.L., Elkassabany, N.M., Kamath, A.F., Liu, J., 2015. Low Albumin Levels, More Than Morbid Obesity, Are Associated With Complications After TKA. Clinical Orthopaedics and Related Research® 473, 3163–3172. URL: http://link.springer.com/10.1007/s11999-015-4333-7, doi:10.1007/s11999-015-4333-7.

[4] Rogers, R.G., Lawrence, E.M., Hummer, R.A., Tilstra, A.M., 2017. Racial/Ethnic Differences in Early-Life Mortality in the United States. Biodemography and Social Biology 63, 189–205. URL: https://www.tandfonline.com/doi/full/10.1080/19485565.2017.1281100, doi:10.1080/19485565.2017.1281100.

5 Appendix

A Descriptive Statistics

History of CVD	No	Yes	Overall
	(N=738)	(N=1080)	(N=1979)
Serum Albumin			
Mean (SD)	3.79 (0.478)	3.67 (0.496)	3.72 (0.501)
Median [Min, Max]	3.80 [1.50, 6.10]	3.70 [0.600, 6.70]	3.80 [0.600, 7.00]
Low Serum Albumin			
No	554 (75.1%)	708 (65.6%)	1374 (69.4%)
Yes	184 (24.9%)	372 (34.4%)	605 (30.6%)
Age			
Mean (SD)	54.2 (16.4)	64.1 (13.4)	60.3 (15.3)
Median [Min, Max]	54.0 [17.0, 90.0]	66.0 [21.0, 93.0]	63.0 [17.0, 93.0]
Sex			
Male	385 (52.2%)	537 (49.7%)	1013 (51.2%)
Female	353 (47.8%)	543 (50.3%)	966 (48.8%)
Smoking Group			
Never	422 (57.2%)	516 (47.8%)	999 (50.5%)
Former	149 (20.2%)	334 (30.9%)	527 (26.6%)
Current	104 (14.1%)	153 (14.2%)	276 (13.9%)
Missing	63 (8.5%)	77 (7.1%)	177 (8.9%)
BMI			
Mean (SD)	24.9 (5.84)	24.6 (5.72)	24.7 (5.76)
Median [Min, Max]	23.8 [12.1, 63.9]	23.7 [13.4, 71.7]	23.7 [12.1, 71.7]
Missing	15 (2.0%)	17 (1.6%)	41 (2.1%)
Total ESRD time			
Mean (SD)	3.42 (3.95)	2.73 (3.12)	2.98 (3.52)
Median [Min, Max]	2.02 [0, 20.2]	1.74 [0, 18.5]	1.80 [0, 20.2]
Undernourished			
No	609 (82.5%)	803 (74.4%)	1515 (76.6%)
Yes	90 (12.2%)	220 (20.4%)	344 (17.4%)
Missing	39 (5.3%)	57 (5.3%)	120 (6.1%)
Race Group			
Caucasian	341 (46.2%)	601 (55.6%)	1033 (52.2%)
African American	327 (44.3%)	384 (35.6%)	766 (38.7%)
Other	57 (7.7%)	79 (7.3%)	150 (7.6%)
Missing	13 (1.8%)	16 (1.5%)	30 (1.5%)
Diabetes			
No	552 (74.8%)	620 (57.4%)	1260 (63.7%)
Yes	185 (25.1%)	460 (42.6%)	710 (35.9%)
Missing	1 (0.1%)	0 (0%)	9 (0.5%)

Table 4: Descriptive statistics of the patient's characteristics in the sample, stratified by CVD. For discrete variables, we provide the count and percentage. We provide the mean, standard deviation, minimum, median, and max values for continuous variables.

	Not Died	Died	Overall
	(N=1489)	(N=490)	(N=1979)
Serum Albumin			
Mean (SD)	3.78 (0.489)	3.56 (0.499)	3.72 (0.501)
Median [Min, Max]	3.80 [0.600, 7.00]	3.60 [1.60, 6.70]	3.80 [0.600, 7.00]
Low Albumin			
No	1107 (74.3%)	267 (54.5%)	1374 (69.4%)
Yes	382 (25.7%)	223 (45.5%)	605 (30.6%)
History of CVD			
No	648 (43.5%)	90 (18.4%)	738 (37.3%)
Yes	741 (49.8%)	339 (69.2%)	1080 (54.6%)
Missing	100 (6.7%)	61 (12.4%)	161 (8.1%)
Age			
Mean (SD)	57.9 (15.4)	67.8 (12.6)	60.3 (15.3)
Median [Min, Max]	60.0 [17.0, 90.0]	69.0 [24.0, 93.0]	63.0 [17.0, 93.0]
Sex			
Male	763 (51.2%)	250 (51.0%)	1013 (51.2%)
Female	726 (48.8%)	240 (49.0%)	966 (48.8%)
Smoking Group			
Never	781 (52.5%)	218 (44.5%)	999 (50.5%)
Former	382 (25.7%)	145 (29.6%)	527 (26.6%)
Current	200 (13.4%)	76 (15.5%)	276 (13.9%)
Missing	126 (8.5%)	51 (10.4%)	177 (8.9%)
BMI			
Mean (SD)	25.1 (5.73)	23.5 (5.68)	24.7 (5.76)
Median [Min, Max]	24.2 [12.1, 63.9]	22.5 [13.4, 71.7]	23.7 [12.1, 71.7]
Missing	21 (1.4%)	20 (4.1%)	41 (2.1%)
Total ESRD time			
Mean (SD)	3.15 (3.70)	2.46 (2.83)	2.98 (3.52)
Median [Min, Max]	1.88 [0, 20.2]	1.65 [0, 18.5]	1.80 [0, 20.2]
Undernourished			
No	1213 (81.5%)	302 (61.6%)	1515 (76.6%)
Yes	183 (12.3%)	161 (32.9%)	344 (17.4%)
Missing	93 (6.2%)	27 (5.5%)	120 (6.1%)
Race Group			
Caucasian	720 (48.4%)	313 (63.9%)	1033 (52.2%)
African American	625 (42.0%)	141 (28.8%)	766 (38.7%)
Other	123 (8.3%)	27 (5.5%)	150 (7.6%)
Missing	21 (1.4%)	9 (1.8%)	30 (1.5%)
Diabetes			
No	966 (64.9%)	294 (60.0%)	1260 (63.7%)
Yes	517 (34.7%)	193 (39.4%)	710 (35.9%)
Missing	6 (0.4%)	3 (0.6%)	9 (0.5%)

Table 5: Descriptive statistics of the patient's characteristics in the sample, stratified by mortality. For discrete variables, we provide the count and percentage. We provide the mean, standard deviation, minimum, median, and max values for continuous variables.

	Overall
	(N=1979)
Albumin	
Mean (SD)	3.72 (0.501)
Median [Min, Max]	3.80 [0.600, 7.00]
Low Albumin	
No	1374 (69.4%)
Yes	605 (30.6%)
CVD history	
No	738 (37.3%)
Yes	1080 (54.6%)
Missing	161 (8.1%)
Age	
Mean (SD)	60.3 (15.3)
Median [Min, Max]	63.0 [17.0, 93.0]
Sex	
Male	1013 (51.2%)
Female	966 (48.8%)
Smoke Group	
Never	999 (50.5%)
Former	527 (26.6%)
Current	276 (13.9%)
Missing	177 (8.9%)
BMI	
Mean (SD)	24.7 (5.76)
Median [Min, Max]	23.7 [12.1, 71.7]
Missing	41 (2.1%)
ESRD time	
Mean (SD)	2.98 (3.52)
Median [Min, Max]	1.80 [0, 20.2]
Undernourished	
No	1515 (76.6%)
Yes	344 (17.4%)
Missing	120 (6.1%)
Race Group	
Caucasian	1033 (52.2%)
African American	766 (38.7%)
Other	150 (7.6%)
Missing	30 (1.5%)
Diabetes	
No	1260 (63.7%)
Yes	710 (35.9%)
Missing	9 (0.5%)

Table 6: Descriptive statistics of the patient's characteristics in the sample. For discrete variables, we provide the count and percentage. We provide the mean, standard deviation, minimum, median, and max values for continuous variables.

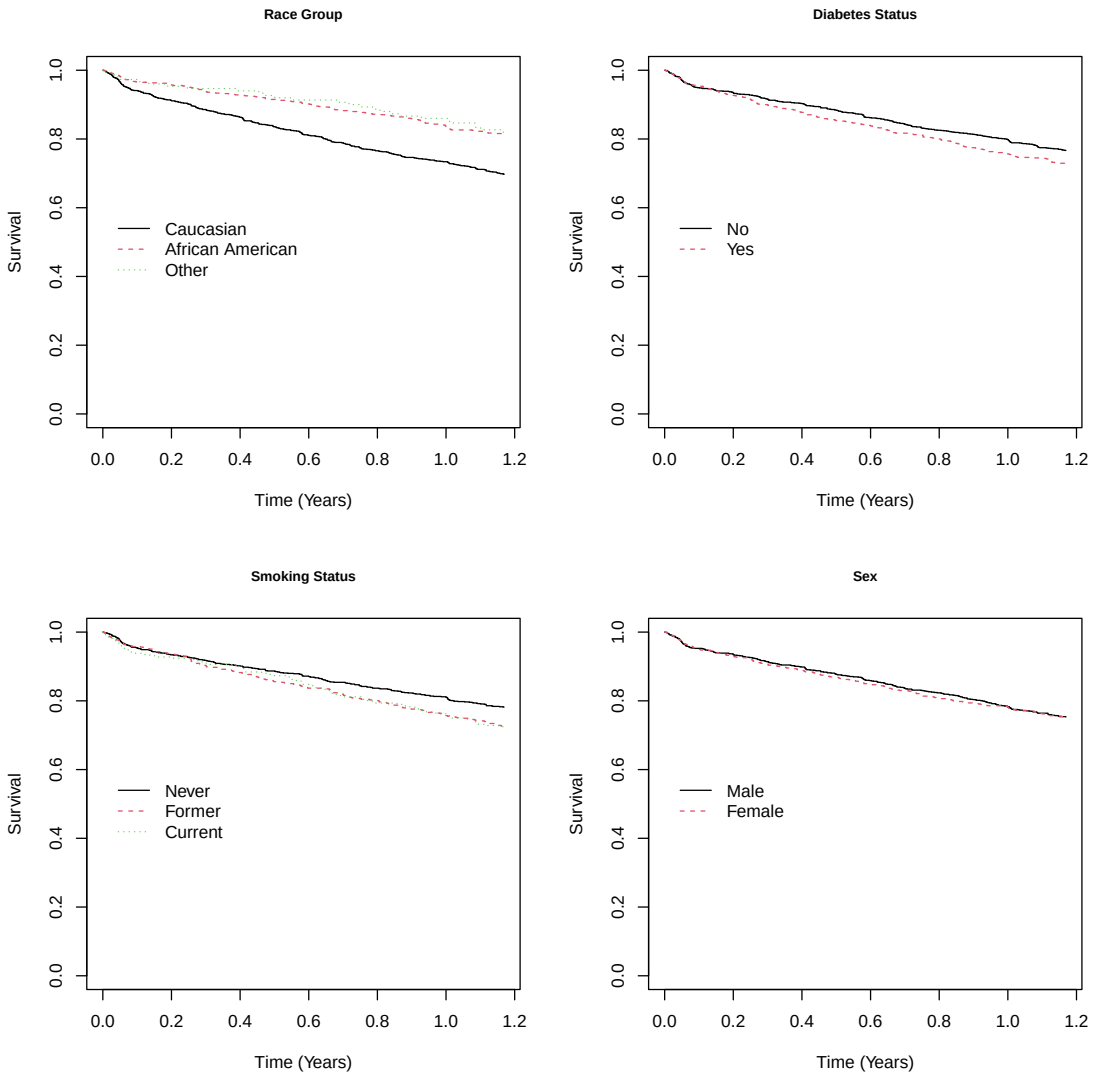

Fig 2: Plot of the survival function to total mortality among hemodialysis patients for different variables.

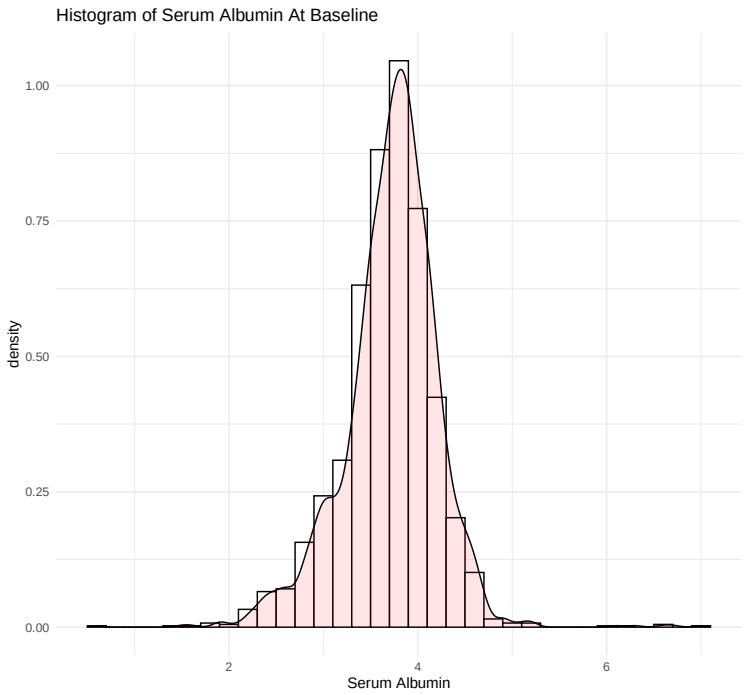

Fig 3: Histogram of albumin at baselines

B Diagnostic

We want to look at any potential outlier, so we plotted the linear predictor vs. deviance residuals plot for our model as shown in Figure 4 and Figure 5. We see a good amount of potential outliers, but since many of the points are marked as outliers, non of the points are outliers, so we should be okay with our analysis.

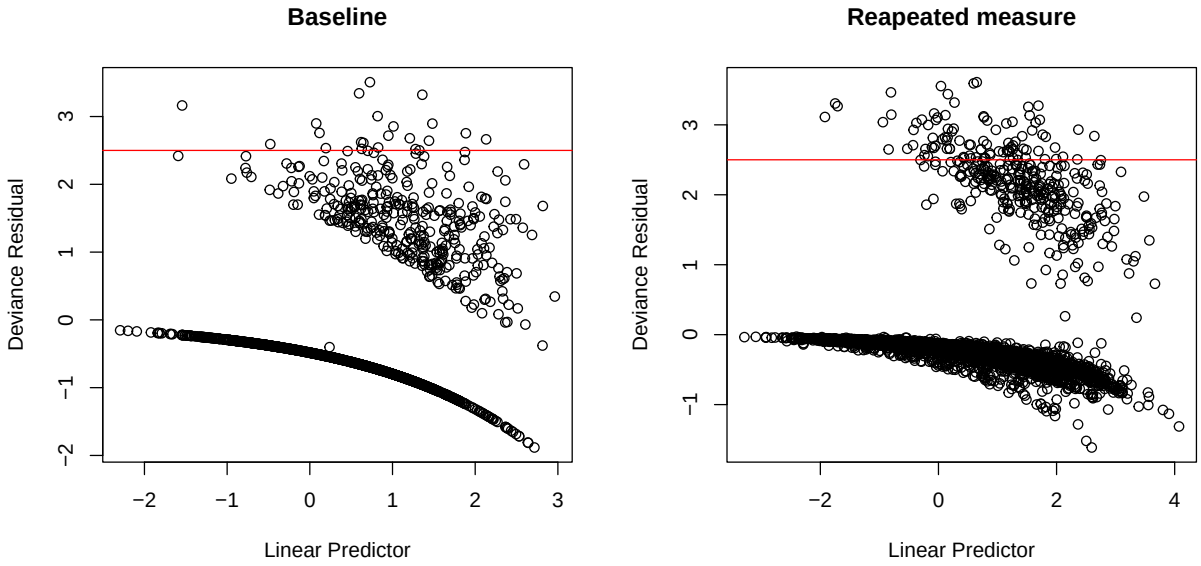

Fig 4: Linear predictor vs Deviance Residual plot for the model discussed in Table 2

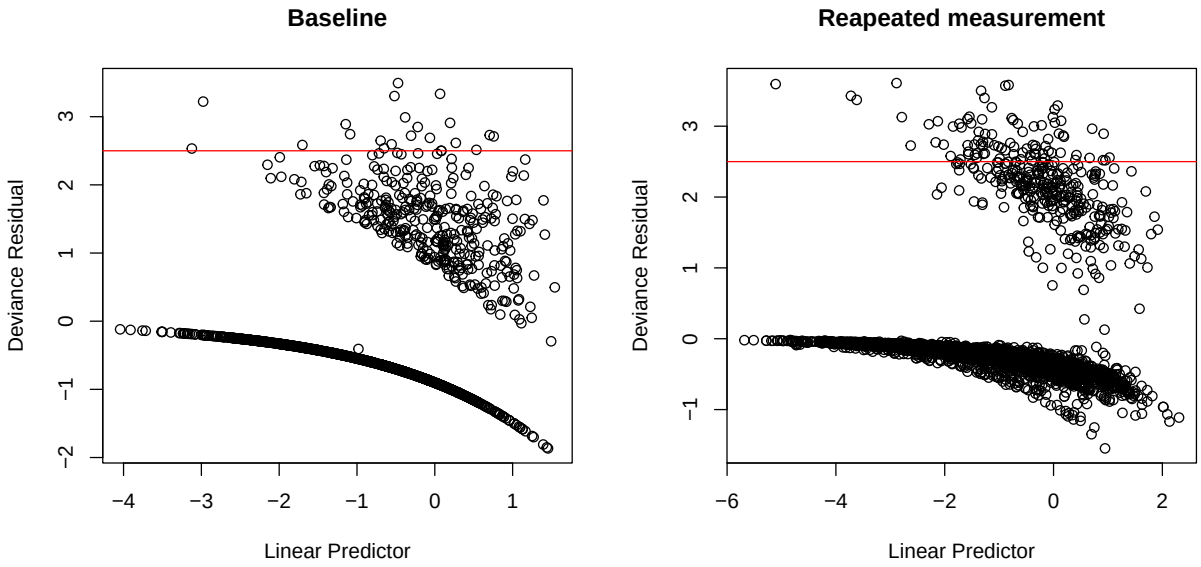

Fig 5: Linear predictor vs Deviance Residual plot for the model discussed in Table 3

C Alternative Sampling Strategy

Lastly, we want to compare a nested case-control design with $M = 4$ control per case and compare it to our baseline model. Table 7 shows hemodialysis patients' estimated relative mortality risk along with their corresponding confidence interval, standard error, and p-value. We see that the baseline model's standard error is lower than the alternative sampling model. We also see similar estimates and that the p-values for both cases are less than 0.001 for serum albumin levels; hence we still have a significants.

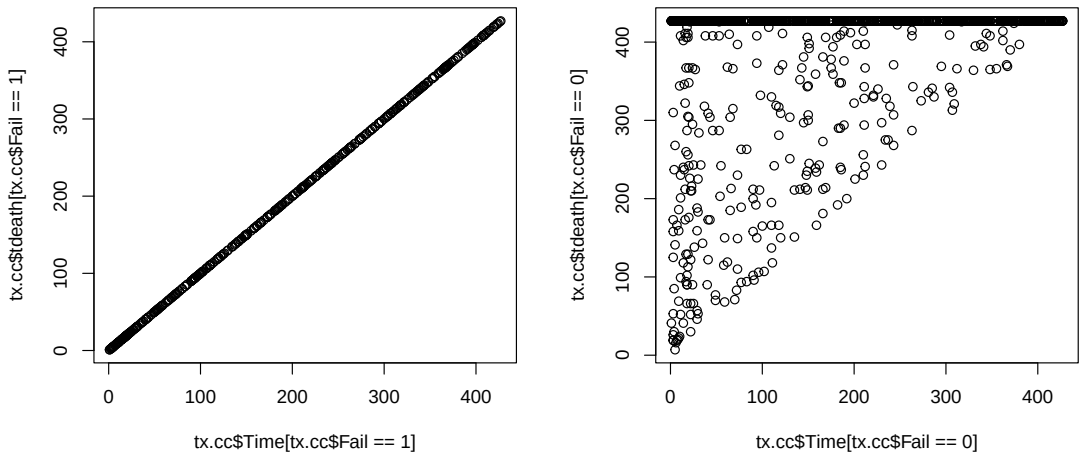

Fig 6: Plot of our nested Case-Control Design with M = 4 controls per case

Covariate	Alternative Sampling			Baseline		
	RR (95% CI)	SE	P-value	RR (95% CI)	SE	P-value
Serum Albumin (g/dL)	0.676 (0.514, 0.891)	0.140	0.0054	0.631 (0.503, 0.792)	0.116	<.0001
< 3.6 g/dL	1.458 (1.118, 1.900)	0.135	0.0054	1.573 (1.261, 1.963)	0.113	<.0001
Cardiovascular Hist.						
No	Reference			Reference		
Yes	2.006 (1.473, 2.733)	0.158	<.0001	2.092 (1.599, 2.738)	0.137	<.0001
Age(10 years)	1.409 (1.265, 1.571)	0.055	<.0001	1.413 (1.291, 1.546)	0.046	<.0001
Sex						
Male	Reference			Reference		
Female	0.930 (0.712, 1.214)	0.136	0.5929	0.905 (0.725, 1.129)	0.113	0.3762
Smoking Status						
Never	Reference			Reference		
Former	0.848 (0.625, 1.151)	0.156	0.2906	0.944 (0.735, 1.213)	0.128	0.6528
Current	1.539 (1.048, 2.261)	0.196	0.0279	1.533 (1.130, 2.080)	0.156	0.0061
BMI (5kg/m^2)	0.849 (0.745, 0.967)	0.066	0.0137	0.894 (0.797, 1.002)	0.058	0.0548
ESRD Total Time	1.029 (0.986, 1.074)	0.022	0.1849	1.010 (0.976, 1.045)	0.017	0.5625
Undernourished						
No	Reference			Reference		
Yes	1.892 (1.405, 2.548)	0.152	<.0001	1.886 (1.482, 2.401)	0.123	<.0001
Race Group						
Caucasian	Reference			Reference		
African American	0.803 (0.607, 1.063)	0.143	0.1251	0.811 (0.642, 1.025)	0.119	0.0798
Other	0.964 (0.555, 1.673)	0.281	0.8962	0.766 (0.481, 1.220)	0.237	0.2615
Diabetes Status						
No	Reference			Reference		
Yes	1.407 (1.056, 1.875)	0.146	0.0197	1.246 (0.991, 1.567)	0.117	0.0594

Table 7: The estimated relative risk of mortality among hemodialysis patients. Along with its corresponding 95% confident interval, standard error, and p values. The first column represents our estimates using alternative sampling, while the second represents our estimates at baseline