

```

In [29]: import pandas as pd
from lifelines import CoxPHFitter
from sklearn.impute import SimpleImputer
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

data = pd.read_excel('data1.xlsx') # Load the data from the .xlsx file

# Preprocess the data: Drop any rows with missing values in the columns of interest
data = data.dropna(subset=['Months', 'DEATH', 'AGE', 'SEX', 'CompositeStage', 'LNInvolment'])

# Handle missing values in other columns
imputer = SimpleImputer(strategy='median')
data[['DEATH', 'AGE', 'CompositeStage', 'LNInvolment', 'Comorbidity']] = imputer.fit_transform(data[['DEATH', 'AGE', 'CompositeStage', 'LNInvolment', 'Comorbidity']])

# Standardize the covariates
scaler = StandardScaler()
data[['DEATH', 'AGE', 'CompositeStage', 'LNInvolment', 'Comorbidity']] = scaler.fit_transform(data[['DEATH', 'AGE', 'CompositeStage', 'LNInvolment', 'Comorbidity']])

# Create a new DataFrame with the required columns for the Buckley-James estimator
buckley_james_data = data[['Months', 'DEATH', 'AGE', 'SEX', 'CompositeStage', 'LNInvolment']]

# Fit the Buckley-James model with custom options
cph = CoxPHFitter(penalizer=0.1) # Set the penalizer parameter to control overfitting
cph.fit(buckley_james_data, 'Months', 'DEATH', show_progress=True) # Set the status variable

# Print the estimated coefficients (summary)
print(cph.summary)

# Access other properties of the fitted model (e.g., hazard ratios, p-values)
# For example, to get the hazard ratios:
print(cph.hazard_ratios_)

# Calculate AIC and BIC
n = len(buckley_james_data)
llf = cph.log_likelihood_
k = cph.params_.shape[0]
aic = -2 * llf + 2 * k
bic = -2 * llf + k * np.log(n)

# Print AIC and BIC
print("AIC:", aic)
print("BIC:", bic)

# Make predictions using the fitted model
# For example, to predict the survival probability at a specific time point for new patient data
new_patient_data = pd.DataFrame({'AGE': [90], 'SEX': [1], 'CompositeStage': [2], 'LNInvolment': [0]})
partial_hazard = cph.predict_partial_hazard(new_patient_data)
survival_prob = 1 - cph.baseline_survival_.iloc[-1]

# Plot the survival curve
plt.plot(cph.baseline_survival_.index, survival_prob.values)
plt.xlabel('Time')
plt.ylabel('Survival Probability')
plt.title('Survival Curve')
plt.show()

# Perform other analyses or visualizations as needed

```

Iteration 1: norm\_delta = 0.66384, step\_size = 0.9500, log\_lik = -1663.17959, new ton\_decrement = 46.04648, seconds\_since\_start = 0.0  
Iteration 2: norm\_delta = 0.03630, step\_size = 0.9500, log\_lik = -1620.53093, new ton\_decrement = 0.19362, seconds\_since\_start = 0.0  
Iteration 3: norm\_delta = 0.00176, step\_size = 0.9500, log\_lik = -1620.33817, new ton\_decrement = 0.00043, seconds\_since\_start = 0.0  
Iteration 4: norm\_delta = 0.00000, step\_size = 1.0000, log\_lik = -1620.33774, new ton\_decrement = 0.00000, seconds\_since\_start = 0.0  
Convergence success after 4 iterations.

	coef	exp(coef)	se(coef)	coef lower 95%	\
covariate					
AGE	0.019975	1.020175	0.055896	-0.089580	
SEX	0.027013	1.027381	0.106745	-0.182203	
CompositeStage	0.531571	1.701603	0.061434	0.411162	
LNInvolment	-0.275748	0.759004	0.053051	-0.379725	
Comorbidity	-0.034023	0.966549	0.054884	-0.141594	
FamiliyHistoryOfCancer	0.003465	1.003471	0.156806	-0.303870	

	coef	upper 95%	exp(coef)	lower 95%	\
covariate					
AGE	0.129529		0.914315		
SEX	0.236229		0.833432		
CompositeStage	0.651980		1.508570		
LNInvolment	-0.171771		0.684049		
Comorbidity	0.073548		0.867974		
FamiliyHistoryOfCancer	0.310800		0.737957		

	exp(coef)	upper 95%	cmp to	z	p	\
covariate						
AGE	1.138292	0.0	0.357349	7.208303e-01		
SEX	1.266464	0.0	0.253064	8.002191e-01		
CompositeStage	1.919337	0.0	8.652682	5.030319e-18		
LNInvolment	0.842172	0.0	-5.197833	2.016254e-07		
Comorbidity	1.076320	0.0	-0.619903	5.353217e-01		
FamiliyHistoryOfCancer	1.364517	0.0	0.022100	9.823684e-01		

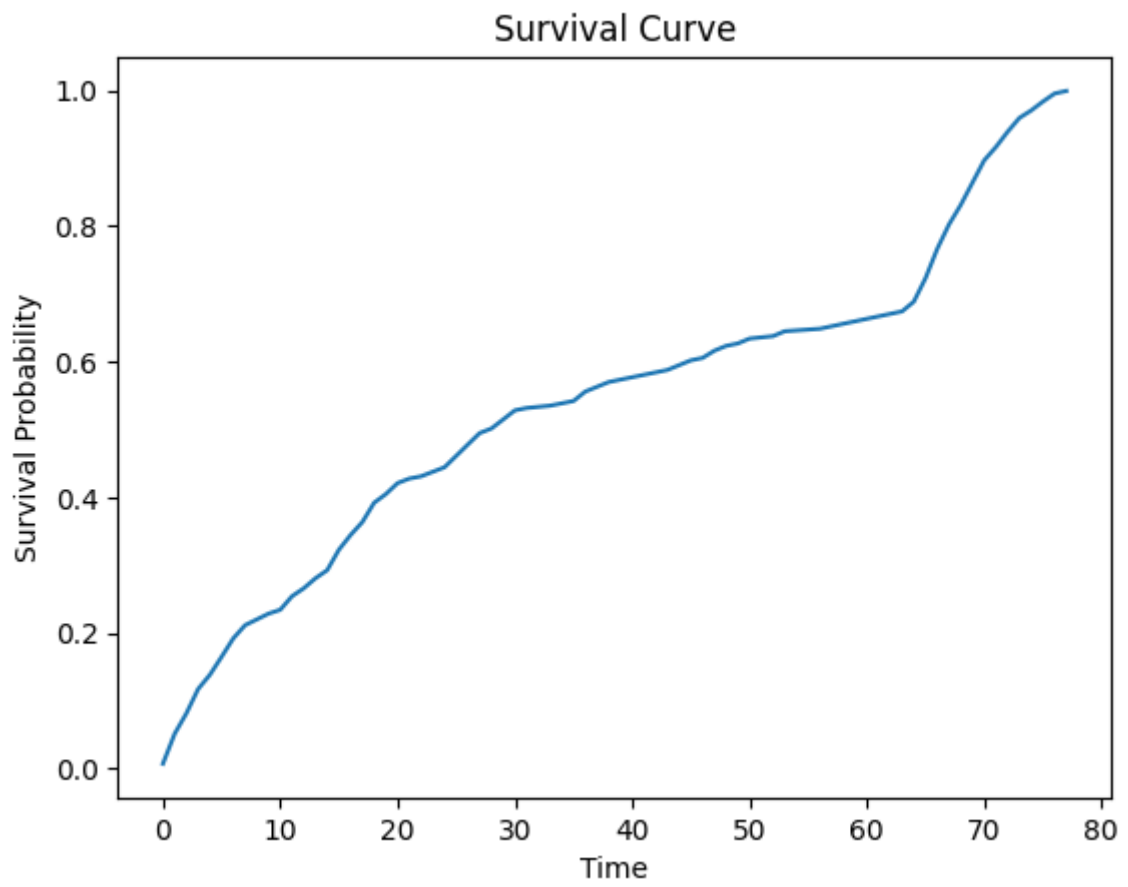
	-log2(p)
covariate	
AGE	0.472268
SEX	0.321533
CompositeStage	57.464056
LNInvolment	22.241820
Comorbidity	0.901522
FamiliyHistoryOfCancer	0.025664

covariate	
AGE	1.020175
SEX	1.027381
CompositeStage	1.701603
LNInvolment	0.759004
Comorbidity	0.966549
FamiliyHistoryOfCancer	1.003471

Name: exp(coef), dtype: float64

AIC: 3252.6754729222544

BIC: 3275.70185560525



In [ ]: