

```

In [7]: import pandas as pd
import numpy as np
from lifelines import CoxPHFitter
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt
from sklearn.impute import SimpleImputer

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

# Define categorical variables
categorical_cols = ['SEX', 'CompositeStage', 'LNInvolment', 'Comorbidity', 'Fami
data[categorical_cols] = data[categorical_cols].astype('category')

# One-hot encode categorical variables
data_encoded = pd.get_dummies(data, columns=categorical_cols, drop_first=True)

# Standardize the covariates
scaler = StandardScaler()
data_encoded[['DEATH', 'AGE']] = scaler.fit_transform(data_encoded[['DEATH', 'AGE']])
buckley_james_data = data_encoded[['Months', 'DEATH', 'AGE']] + [col for col in c

data = data.dropna(subset=['Months', 'DEATH', 'AGE', 'SEX', 'CompositeStage', 'L

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

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

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

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

# Print the estimated coefficients (summary)
print(cph.summary)
# Perform univariate analysis
univariate_results = []
univariate_aic_bic = []
for col in buckley_james_data.columns:
    if col not in ['Months', 'DEATH', 'AGE']:
        cph_univariate = CoxPHFitter(penalizer=0.1)
        cph_univariate.fit(buckley_james_data[[col, 'Months', 'DEATH', 'AGE']],
        p_value = cph_univariate.summary['p'][col]
        univariate_results.append((col, p_value))
        n = len(buckley_james_data)
        llf = cph_univariate.log_likelihood_
        k = cph_univariate.params_.shape[0]
        aic = -2 * llf + 2 * k
        bic = -2 * llf + k * np.log(n)
        univariate_aic_bic.append((col, aic, bic))
        print(f"AIC value of {col}:", aic)
        print(f"BIC value of {col}:", bic)

```

```

# Select significant variables
significant_variables = [(var, p_value) for var, p_value in univariate_results.items() if p_value < 0.05]
print("Significant variables from univariate analysis:")
for var, p_value in significant_variables:
    print(f"\n{var}: p-value={p_value}")

# Fit the multivariate model
cph_multivariate = CoxPHFitter(penalizer=0.1)
cph_multivariate.fit(buckley_james_data[['Months', 'DEATH', 'AGE'] + [var for var in significant_variables[0][0]]))

# Calculate AIC and BIC for multivariate model
n = len(buckley_james_data)
llf = cph_multivariate.log_likelihood_
k = cph_multivariate.params_.shape[0]
multivariate_aic = -2 * llf + 2 * k
multivariate_bic = -2 * llf + k * np.log(n)

# Print AIC and BIC for multivariate model
print("AIC value of the multivariate model:", multivariate_aic)
print("BIC value of the multivariate model:", multivariate_bic)

# Print AIC and BIC for univariate models
print("AIC and BIC for univariate models:")
for col, aic, bic in univariate_aic_bic:
    print(f"{col}: AIC={aic}, BIC={bic}")

```

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	
FamilyHistoryOfCancer	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		
FamilyHistoryOfCancer	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		
FamilyHistoryOfCancer	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
FamilyHistoryOfCancer	0.025664

Iteration 1: norm_delta = 0.02554, step_size = 0.9500, log_lik = -1663.17959, new ton_decrement = 0.12225, seconds_since_start = 0.0
 Iteration 2: norm_delta = 0.00120, step_size = 0.9500, log_lik = -1663.05786, new ton_decrement = 0.00027, seconds_since_start = 0.0
 Iteration 3: norm_delta = 0.00006, step_size = 0.9500, log_lik = -1663.05758, new ton_decrement = 0.00000, seconds_since_start = 0.0
 Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1663.05758, new ton_decrement = 0.00000, seconds_since_start = 0.0
 Convergence success after 4 iterations.

AIC value of SEX: 3330.115165027162

BIC value of SEX: 3337.790625921494

Iteration 1: norm_delta = 0.43180, step_size = 0.9500, log_lik = -1663.17959, new ton_decrement = 27.14843, seconds_since_start = 0.0
 Iteration 2: norm_delta = 0.04180, step_size = 0.9500, log_lik = -1635.50777, new ton_decrement = 0.23026, seconds_since_start = 0.0
 Iteration 3: norm_delta = 0.00240, step_size = 0.9500, log_lik = -1635.27713, new

ton_decrement = 0.00075, seconds_since_start = 0.0
 Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1635.27638, new
 ton_decrement = 0.00000, seconds_since_start = 0.0
 Convergence success after 4 iterations.
 AIC value of CompositeStage: 3274.552767693642
 BIC value of CompositeStage: 3282.228228587974
 Iteration 1: norm_delta = 0.14048, step_size = 0.9500, log_lik = -1663.17959, new
 ton_decrement = 4.00489, seconds_since_start = 0.0
 Iteration 2: norm_delta = 0.01332, step_size = 0.9500, log_lik = -1659.09305, new
 ton_decrement = 0.03361, seconds_since_start = 0.0
 Iteration 3: norm_delta = 0.00074, step_size = 0.9500, log_lik = -1659.05942, new
 ton_decrement = 0.00010, seconds_since_start = 0.0
 Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1659.05932, new
 ton_decrement = 0.00000, seconds_since_start = 0.0
 Convergence success after 4 iterations.
 AIC value of LNInvolment: 3322.1186323122874
 BIC value of LNInvolment: 3329.7940932066194
 Iteration 1: norm_delta = 0.06728, step_size = 0.9500, log_lik = -1663.17959, new
 ton_decrement = 0.79933, seconds_since_start = 0.0
 Iteration 2: norm_delta = 0.00280, step_size = 0.9500, log_lik = -1662.38626, new
 ton_decrement = 0.00141, seconds_since_start = 0.0
 Iteration 3: norm_delta = 0.00014, step_size = 0.9500, log_lik = -1662.38485, new
 ton_decrement = 0.00000, seconds_since_start = 0.0
 Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1662.38484, new
 ton_decrement = 0.00000, seconds_since_start = 0.0
 Convergence success after 4 iterations.
 AIC value of Comorbidity: 3328.7696858848362
 BIC value of Comorbidity: 3336.445146779168
 Iteration 1: norm_delta = 0.03241, step_size = 0.9500, log_lik = -1663.17959, new
 ton_decrement = 0.19925, seconds_since_start = 0.0
 Iteration 2: norm_delta = 0.00091, step_size = 0.9500, log_lik = -1662.98363, new
 ton_decrement = 0.00017, seconds_since_start = 0.0
 Iteration 3: norm_delta = 0.00005, step_size = 0.9500, log_lik = -1662.98346, new
 ton_decrement = 0.00000, seconds_since_start = 0.0
 Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1662.98346, new
 ton_decrement = 0.00000, seconds_since_start = 0.0
 Convergence success after 4 iterations.
 AIC value of FamilyHistoryOfCancer: 3329.9669227876907
 BIC value of FamilyHistoryOfCancer: 3337.6423836820227
 Significant variables from univariate analysis:

CompositeStage: p-value=3.1332497028365436e-13

LNInvolment: p-value=0.005178096993930303
 Iteration 1: norm_delta = 0.66556, step_size = 0.9500, log_lik = -1663.17959, new
 ton_decrement = 45.81771, seconds_since_start = 0.0
 Iteration 2: norm_delta = 0.03533, step_size = 0.9500, log_lik = -1620.74129, new
 ton_decrement = 0.18228, seconds_since_start = 0.0
 Iteration 3: norm_delta = 0.00167, step_size = 0.9500, log_lik = -1620.55993, new
 ton_decrement = 0.00039, seconds_since_start = 0.0
 Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1620.55954, new
 ton_decrement = 0.00000, seconds_since_start = 0.1
 Convergence success after 4 iterations.
 AIC value of the multivariate model: 3247.1190700944935
 BIC value of the multivariate model: 3258.6322614359915
 AIC and BIC for univariate models:
 SEX: AIC=3330.115165027162, BIC=3337.790625921494
 CompositeStage: AIC=3274.552767693642, BIC=3282.228228587974
 LNInvolment: AIC=3322.1186323122874, BIC=3329.7940932066194

Comorbidity: AIC=3328.7696858848362, BIC=3336.445146779168
FamilyHistoryOfCancer: AIC=3329.9669227876907, BIC=3337.6423836820227

In []: