In [1]:

```
import pandas as pd
import numpy as np
from lifelines import CoxPHFitter
from sklearn.preprocessing import StandardScaler
import matplotlib.pyplot as plt

# 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', 'AG
buckley_james_data = data_encoded[['Months', 'DEATH', 'AGE'] + [col for col in c

# 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 i
print("Significant variables from univariate analysis:")
for var, p_value in significant_variables:
    print(f"{var}: significant-value={p_value}")

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

# 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 = 0.02554, 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_2: 3330.115165027162
BIC value of SEX_2: 3337.790625921494
Iteration 1: norm_delta = 0.14622, step_size = 0.9500, log_lik = -1663.17959, new
ton_decrement = 4.31318, seconds_since_start = 0.0
Iteration 2: norm_delta = 0.01327, step_size = 0.9500, log_lik = -1658.78794, new
ton_decrement = 0.03279, seconds_since_start = 0.0
Iteration 3: norm_delta = 0.00073, step_size = 0.9500, log_lik = -1658.75515, new
ton_decrement = 0.00010, seconds_since_start = 0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1658.75505, new
ton_decrement = 0.00000, seconds_since_start = 0.0
Convergence success after 4 iterations.
AIC value of CompositeStage_2: 3321.5101024525675
BIC value of CompositeStage_2: 3329.1855633468995
Iteration 1: norm_delta = 0.17583, step_size = 0.9500, log_lik = -1663.17959, new
ton_decrement = 6.44524, seconds_since_start = 0.0
Iteration 2: norm_delta = 0.01738, step_size = 0.9500, log_lik = -1656.59487, new
ton_decrement = 0.05734, seconds_since_start = 0.0
Iteration 3: norm_delta = 0.00098, step_size = 0.9500, log_lik = -1656.53746, new
ton_decrement = 0.00018, seconds_since_start = 0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1656.53728, new
ton_decrement = 0.00000, seconds_since_start = 0.0
Convergence success after 4 iterations.
AIC value of CompositeStage_3: 3317.0745678980034
BIC value of CompositeStage_3: 3324.7500287923353
Iteration 1: norm_delta = 0.70165, step_size = 0.9500, log_lik = -1663.17959, new
ton_decrement = 50.56989, seconds_since_start = 0.0
Iteration 2: norm_delta = 0.08686, step_size = 0.9500, log_lik = -1620.40900, new
ton_decrement = 1.08200, seconds_since_start = 0.0
Iteration 3: norm_delta = 0.00480, step_size = 0.9500, log_lik = -1619.32672, new
ton_decrement = 0.00327, seconds_since_start = 0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1619.32345, new
ton_decrement = 0.00000, seconds_since_start = 0.0
Convergence success after 4 iterations.
AIC value of CompositeStage_4: 3242.6468994706224
BIC value of CompositeStage_4: 3250.3223603649544
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 = 0.1
Convergence success after 4 iterations.
AIC value of LNInvolment_1: 3322.1186323122874
BIC value of LNInvolment_1: 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_1: 3328.7696858848362
BIC value of Comorbidity_1: 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 FamiliyHistoryOfCancer_1: 3329.9669227876907
BIC value of FamiliyHistoryOfCancer_1: 3337.6423836820227
Significant variables from univariate analysis:
CompositeStage_2: significant-value=0.0036709671779481512
CompositeStage_3: significant-value=0.00038245728111221104
CompositeStage_4: significant-value=3.861516598671974e-22
LNInvolment_1: significant-value=0.005178096993930311
Iteration 1: norm_delta = 0.68301, step_size = 0.9500, log_lik = -1663.17959, new
ton_decrement = 54.30464, seconds_since_start = 0.0
Iteration 2: norm_delta = 0.08057, step_size = 0.9500, log_lik = -1616.92886, new
ton_decrement = 1.14381, seconds_since_start = 0.0
Iteration 3: norm_delta = 0.00450, step_size = 0.9500, log_lik = -1615.78591, new
ton_decrement = 0.00329, seconds_since_start = 0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1615.78262, new
ton_decrement = 0.00000, seconds_since_start = 0.0
Convergence success after 4 iterations.
AIC value of the multivariate model: 3241.5652333399553
BIC value of the multivariate model: 3260.753885575785
AIC and BIC for univariate models:
SEX_2: AIC=3330.115165027162, BIC=3337.790625921494
CompositeStage_2: AIC=3321.5101024525675, BIC=3329.1855633468995
CompositeStage_3: AIC=3317.0745678980034, BIC=3324.7500287923353
CompositeStage_4: AIC=3242.6468994706224, BIC=3250.3223603649544
LNInvolment_1: AIC=3322.1186323122874, BIC=3329.7940932066194
Comorbidity_1: AIC=3328.7696858848362, BIC=3336.445146779168
FamiliyHistoryOfCancer_1: AIC=3329.9669227876907, BIC=3337.6423836820227
```

In [ ]: