In [18]:

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

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

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

# Create a dictionary to store the AIC_partial and BIC values for each variable
aic_bic_results = {}

# Perform univariate analysis for each column separately
for col in data.columns:
    if col not in ['Months', 'DEATH', 'AGE']:
        cph_univariate = CoxPHFitter(penalizer=0.1)
        cph_univariate.fit(data[[col, 'Months', 'DEATH', 'AGE']], 'Months', 'DEATH', show_progress=True)

        # Calculate AIC_partial for the current column
        aic_partial = cph_univariate.AIC_partial_

        # Calculate the number of parameters
        num_params = cph_univariate.params_.shape[0]

        # Calculate BIC using the formula: BIC = -2 * log-likelihood + num_params * log(n)
        n = data.shape[0]
        log_likelihood = cph_univariate.log_likelihood_
        bic = -2 * log_likelihood + num_params * np.log(n)

        # Store the AIC_partial and BIC values in the dictionary
        aic_bic_results[col] = {'AIC_partial': aic_partial, 'BIC': bic}

        # Display the univariate analysis results for the current column
        print(f"\nUnivariate analysis for column: {col}")
        print(cph_univariate.summary)
```

```
print("\nAIC and BIC results:")
for col, values in aic_bic_results.items():
    print(f"Variable: {col}")
    print(f"AIC: {values['AIC_partial']}")
    print(f"BIC: {values['BIC']}")
```

```
Iteration 1: norm_delta = 0.24443, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 12.49691, seconds_since_start
= 0.0
Iteration 2: norm_delta = 0.00668, step_size = 0.9500, log_lik = -1650.99935, newton_decrement = 0.00772, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00033, step_size = 0.9500, log_lik = -1650.99166, newton_decrement = 0.00002, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1650.99164, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.

Univariate analysis for column: ID
               coef  exp(coef)  se(coef)  coef lower 95%  coef upper 95%  \
covariate
ID         0.002337   1.002340  0.000471        0.001413        0.003261
AGE       -0.048967   0.952212  0.052646       -0.152151        0.054217

           exp(coef) lower 95%  exp(coef) upper 95%  cmp to         z  \
covariate
ID                    1.001414             1.003267     0.0  4.957504
AGE                   0.858858             1.055713     0.0 -0.930126

                      p  -log2(p)
covariate
ID         7.140460e-07  20.41748
AGE        3.523058e-01   1.50510
Iteration 1: norm_delta = 0.02554, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 0.12225, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00120, step_size = 0.9500, log_lik = -1663.05786, newton_decrement = 0.00027, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00006, step_size = 0.9500, log_lik = -1663.05758, newton_decrement = 0.00000, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1663.05758, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.

Univariate analysis for column: SEX
               coef  exp(coef)  se(coef)  coef lower 95%  coef upper 95%  \
covariate
SEX        0.036028   1.036685  0.105515       -0.170776        0.242833
AGE       -0.018326   0.981841  0.052213       -0.120660        0.084009
```

```
           exp(coef) lower 95%  exp(coef) upper 95%  cmp to         z  \
covariate
SEX                   0.843010             1.274856     0.0  0.341454
AGE                   0.886335             1.087639     0.0 -0.350979

                  p  -log2(p)
covariate
SEX        0.732762  0.448583
AGE        0.725604  0.462746
Iteration 1: norm_delta = 0.43180, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 27.14843, seconds_since_start
= 0.0
Iteration 2: norm_delta = 0.04180, step_size = 0.9500, log_lik = -1635.50777, newton_decrement = 0.23026, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00240, step_size = 0.9500, log_lik = -1635.27713, newton_decrement = 0.00075, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1635.27638, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.

Univariate analysis for column: CompositeStage
                    coef  exp(coef)  se(coef)  coef lower 95%  coef upper 95%  \
covariate
CompositeStage  0.451465   1.570611  0.061942        0.330061        0.572868
AGE             0.013370   1.013460  0.053478       -0.091446        0.118185

                exp(coef) lower 95%  exp(coef) upper 95%  cmp to         z  \
covariate
CompositeStage             1.391053             1.773346     0.0  7.288543
AGE                        0.912611             1.125453     0.0  0.250006

                           p   -log2(p)
covariate
CompositeStage  3.133250e-13  41.537405
AGE             8.025824e-01   0.317279
Iteration 1: norm_delta = 0.14048, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 4.00489, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.01332, step_size = 0.9500, log_lik = -1659.09305, newton_decrement = 0.03361, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00074, step_size = 0.9500, log_lik = -1659.05942, newton_decrement = 0.00010, seconds_since_start =
```

```
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1659.05932, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.

Univariate analysis for column: LNInvolment
                coef  exp(coef)  se(coef)  coef lower 95%  coef upper 95%  \
covariate
LNInvolment -0.143911   0.865965  0.051475       -0.244800       -0.043022
AGE         -0.027609   0.972769  0.052098       -0.129719        0.074502

             exp(coef) lower 95%  exp(coef) upper 95%  cmp to         z  \
covariate
LNInvolment              0.782861             0.957891     0.0 -2.795740
AGE                      0.878342             1.077347     0.0 -0.529935

                   p  -log2(p)
covariate
LNInvolment  0.005178  7.593362
AGE          0.596157  0.746236
Iteration 1: norm_delta = 0.06728, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 0.79933, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00280, step_size = 0.9500, log_lik = -1662.38626, newton_decrement = 0.00141, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00014, step_size = 0.9500, log_lik = -1662.38485, newton_decrement = 0.00000, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1662.38484, newton_decrement = 0.00000, seconds_since_start =
0.1
Convergence success after 4 iterations.

Univariate analysis for column: Comorbidity
                coef  exp(coef)  se(coef)  coef lower 95%  coef upper 95%  \
covariate
Comorbidity -0.066494   0.935669  0.054836       -0.173970        0.040982
AGE          0.004068   1.004076  0.055230       -0.104182        0.112317

             exp(coef) lower 95%  exp(coef) upper 95%  cmp to         z  \
covariate
Comorbidity              0.840322             1.041833     0.0 -1.212605
AGE                      0.901062             1.118867     0.0  0.073650
```

```
                    p  -log2(p)
covariate
Comorbidity  0.225281  2.150203
AGE          0.941289  0.087290
Iteration 1: norm_delta = 0.03241, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 0.19925, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00091, step_size = 0.9500, log_lik = -1662.98363, newton_decrement = 0.00017, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00005, step_size = 0.9500, log_lik = -1662.98346, newton_decrement = 0.00000, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1662.98346, newton_decrement = 0.00000, seconds_since_start =
0.1
Convergence success after 4 iterations.

Univariate analysis for column: FamiliyHistoryOfCancer
                            coef  exp(coef)  se(coef)  coef lower 95%  \
covariate
FamiliyHistoryOfCancer  0.080913   1.084277  0.155784       -0.224418
AGE                    -0.016309   0.983823  0.052481       -0.119169

                        coef upper 95%  exp(coef) lower 95%  \
covariate
FamiliyHistoryOfCancer        0.386244             0.798981
AGE                           0.086551             0.887658

                        exp(coef) upper 95%  cmp to         z         p  \
covariate
FamiliyHistoryOfCancer             1.471444     0.0  0.519395  0.603485
AGE                                1.090407     0.0 -0.310763  0.755981

                        -log2(p)
covariate
FamiliyHistoryOfCancer  0.728609
AGE                     0.403579

AIC and BIC results:
Variable: ID
AIC: 3305.983279633574
BIC: 3313.658740527906
```

```
Variable: SEX
AIC: 3330.115165027162
BIC: 3337.790625921494
Variable: CompositeStage
AIC: 3274.552767693642
BIC: 3282.228228587974
Variable: LNInvolment
AIC: 3322.1186323122874
BIC: 3329.7940932066194
Variable: Comorbidity
AIC: 3328.7696858848362
BIC: 3336.445146779168
Variable: FamiliyHistoryOfCancer
AIC: 3329.9669227876907
BIC: 3337.6423836820227
```

In [17]:
```python
# Determine the significant variables based on the p-values
significance_level = 0.05
significant_variables = [col for col, p_value in p_values.items() if p_value < significance_level]

# Display the significant variables
print("\nSignificant variables:")
for variable in significant_variables:
    print(variable)
```

```
Significant variables:
ID
CompositeStage
LNInvolment
```

In [12]:
```python
# Determine the significant variables based on the p-values
significance_level = 0.05
significant_variables = [col for col, p_value in p_values.items() if p_value < significance_level]

# Display the values of the significant variables
print("\nValues of significant variables:")
for variable in significant_variables:
    print(f"{variable}:")
    print(data[variable])
    print()
```

```
Values of significant variables:
ID:
0        1
1        2
2        3
3        4
4        5
      ...
338    339
339    340
340    341
341    342
342    343
Name: ID, Length: 343, dtype: int64

CompositeStage:
0      0.032170
1     -2.174702
2     -1.071266
3     -1.071266
4      0.032170
         ...
338    0.032170
339   -1.071266
340   -1.071266
341    1.135606
342    1.135606
Name: CompositeStage, Length: 343, dtype: float64

LNInvolment:
0      1.604031
1     -0.623429
2     -0.623429
3     -0.623429
4      1.604031
         ...
338    1.604031
339   -0.623429
340   -0.623429
341   -0.623429
```

```
342    1.604031
Name: LNInvolment, Length: 343, dtype: float64
```

In [16]:
```python
categorical_variables = {}

# Convert significant variables into categorical variables
for variable in significant_variables:
    unique_values = data[variable].unique()
    categorical_variables[variable] = pd.Categorical(data[variable], categories=unique_values)

# Perform multivariate analysis using Cox proportional hazards model
cph_multivariate = CoxPHFitter(penalizer=0.1)
cph_multivariate.fit(pd.DataFrame(categorical_variables).join(data[['Months', 'DEATH', 'AGE']]), 'Months', 'DEATH', show_progr

# Display the multivariate analysis results
print("\nMultivariate analysis results:")
print(cph_multivariate.summary)
```

```
Iteration 1: norm_delta = 0.62629, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 49.00012, seconds_since_start
= 0.0
Iteration 2: norm_delta = 0.04251, step_size = 0.9500, log_lik = -1617.00189, newton_decrement = 0.21024, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00186, step_size = 0.9500, log_lik = -1616.79304, newton_decrement = 0.00040, seconds_since_start =
0.1
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1616.79265, newton_decrement = 0.00000, seconds_since_start =
0.1
Convergence success after 4 iterations.

Multivariate analysis results:
                    coef  exp(coef)  se(coef)  coef lower 95%  coef upper 95%  \
covariate
ID              0.001434   1.001435  0.000519        0.000417        0.002451
CompositeStage  0.504181   1.655629  0.062374        0.381929        0.626432
LNInvolment    -0.226724   0.797141  0.055932       -0.336350       -0.117098
AGE            -0.012415   0.987662  0.053825       -0.117910        0.093079

                exp(coef) lower 95%  exp(coef) upper 95%  cmp to         z  \
covariate
ID                          1.000417             1.002454     0.0  2.763220
CompositeStage              1.465109             1.870924     0.0  8.083145
LNInvolment                 0.714373             0.889498     0.0 -4.053532
AGE                         0.888776             1.097549     0.0 -0.230658

                            p   -log2(p)
covariate
ID              5.723417e-03   7.448908
CompositeStage  6.311768e-16  50.492805
LNInvolment     5.045008e-05  14.274784
AGE             8.175801e-01   0.290568
```

In [15]:

```python
n = data.shape[0]
log_likelihood = cph_multivariate.log_likelihood_
num_params = cph_multivariate.params_.shape[0]
aic = -2 * log_likelihood + 2 * num_params
bic = -2 * log_likelihood + num_params * np.log(n)

# Display the multivariate analysis results, AIC, and BIC
print("\nMultivariate analysis results:")
print(cph_multivariate.summary)
```

```
print(f"AIC: {aic}")
print(f"BIC: {bic}")
```

```
Multivariate analysis results:
                    coef  exp(coef)  se(coef)  coef lower 95%  coef upper 95%  \
covariate
ID              0.001434   1.001435  0.000519        0.000417        0.002451
CompositeStage  0.504181   1.655629  0.062374        0.381929        0.626432
LNInvolment    -0.226724   0.797141  0.055932       -0.336350       -0.117098
AGE            -0.012415   0.987662  0.053825       -0.117910        0.093079

                exp(coef) lower 95%  exp(coef) upper 95%  cmp to         z  \
covariate
ID                          1.000417             1.002454     0.0  2.763220
CompositeStage              1.465109             1.870924     0.0  8.083145
LNInvolment                 0.714373             0.889498     0.0 -4.053532
AGE                         0.888776             1.097549     0.0 -0.230658

                            p   -log2(p)
covariate
ID              5.723417e-03   7.448908
CompositeStage  6.311768e-16  50.492805
LNInvolment     5.045008e-05  14.274784
AGE             8.175801e-01   0.290568
AIC: 3241.5852962379945
BIC: 3256.9362180266585
```

In [ ]: