In [1]:

```
import pandas as pd
from lifelines import CoxPHFitter
import numpy as np

# Read the Excel file into a pandas DataFrame
data = pd.read_excel('DATA1.xlsx')

# Create a CoxPHFitter object
cph = CoxPHFitter()

# Fit the Cox proportional hazards model
cph.fit(data, duration_col='Months', event_col='DEATH')

# Print the summary of the model
print(cph.summary)
```

```
                           coef  exp(coef)  se(coef)  coef lower 95%  \
covariate
ID                         0.002267   1.002270  0.000807        0.000685
AGE                        0.015419   1.015539  0.007463        0.000791
SEX                       -0.122082   0.885076  0.159489       -0.434675
CompositeStage             0.914415   2.495315  0.103353        0.711846
LNInvolment               -0.685620   0.503778  0.192838       -1.063576
Comorbidity               -0.111984   0.894058  0.164026       -0.433468
FamiliyHistoryOfCancer    -0.274577   0.759893  0.234633       -0.734448

                          coef upper 95%  exp(coef) lower 95%  \
covariate
ID                              0.003849             1.000685
AGE                             0.030047             1.000792
SEX                             0.190512             0.647475
CompositeStage                  1.116984             2.037750
LNInvolment                    -0.307665             0.345219
Comorbidity                     0.209500             0.648257
FamiliyHistoryOfCancer          0.185294             0.479770

                          exp(coef) upper 95%  cmp to         z             p  \
covariate
ID                                   1.003856     0.0  2.809049  4.968812e-03
AGE                                  1.030503     0.0  2.065987  3.882974e-02
SEX                                  1.209868     0.0 -0.765454  4.440015e-01
CompositeStage                       3.055624     0.0  8.847472  8.952438e-19
LNInvolment                          0.735162     0.0 -3.555420  3.773764e-04
Comorbidity                          1.233061     0.0 -0.682725  4.947809e-01
FamiliyHistoryOfCancer               1.203573     0.0 -1.170243  2.419033e-01

                           -log2(p)
covariate
ID                         7.652883
AGE                        4.686694
SEX                        1.171363
CompositeStage            59.954353
LNInvolment               11.371708
Comorbidity                1.015138
FamiliyHistoryOfCancer     2.047498
```

In [4]:
```
univariate_results = []
univariate_aic_bic = []
for col in data.columns:
    if col not in ['Months', 'ID']:
        cph_univariate = CoxPHFitter(penalizer=0.1)
        cph_univariate.fit(data[['Months', 'ID', col]], duration_col='Months', event_col='ID', show_progress=True)
        univariate_results.append((col, cph_univariate.print_summary()))
```

```
Iteration 1: norm_delta = 0.77722, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 70.49318, seconds_since_start
= 0.0
Iteration 2: norm_delta = 0.01332, step_size = 0.9500, log_lik = -1595.37158, newton_decrement = 0.02032, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00069, step_size = 0.9500, log_lik = -1595.35129, newton_decrement = 0.00005, seconds_since_start =
0.1
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1595.35124, newton_decrement = 0.00000, seconds_since_start =
0.1
Convergence success after 4 iterations.
```

model	lifelines.CoxPHFitter
duration col	'Months'
event col	'ID'
penalizer	0.1
l1 ratio	0.0
baseline estimation	breslow
number of observations	343
number of events observed	343
partial log-likelihood	-1595.35
time fit was run	2023-08-14 11:17:33 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
DEATH	1.50	4.49	0.13	1.24	1.76	3.46	5.82	0.00	11.35	<0.005	96.84

Concordance	0.66
Partial AIC	3192.70
log-likelihood ratio test	135.66 on 1 df
-log2(p) of ll-ratio test	101.73

```
Iteration 1: norm_delta = 0.01879, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 0.06380, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00085, step_size = 0.9500, log_lik = -1663.11614, newton_decrement = 0.00013, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00004, step_size = 0.9500, log_lik = -1663.11600, newton_decrement = 0.00000, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1663.11600, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.
```

model	lifelines.CoxPHFitter
duration col	'Months'
event col	'ID'
penalizer	0.1
l1 ratio	0.0
baseline estimation	breslow
number of observations	343
number of events observed	343
partial log-likelihood	-1663.12
time fit was run	2023-08-14 11:17:34 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
AGE	-0.00	1.00	0.00	-0.01	0.01	0.99	1.01	0.00	-0.36	0.72	0.47

Concordance	0.51
Partial AIC	3328.23
log-likelihood ratio test	0.13 on 1 df
-log2(p) of ll-ratio test	0.47

```
Iteration 1: norm_delta = 0.01792, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 0.06049, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00095, step_size = 0.9500, log_lik = -1663.11915, newton_decrement = 0.00017, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00005, step_size = 0.9500, log_lik = -1663.11898, newton_decrement = 0.00000, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1663.11898, newton_decrement = 0.00000, seconds_since_start =
0.1
Convergence success after 4 iterations.
```

model	lifelines.CoxPHFitter
duration col	'Months'
event col	'ID'
penalizer	0.1
l1 ratio	0.0
baseline estimation	breslow
number of observations	343
number of events observed	343
partial log-likelihood	-1663.12
time fit was run	2023-08-14 11:17:34 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
SEX	0.04	1.04	0.11	-0.17	0.24	0.84	1.28	0.00	0.35	0.73	0.46

Concordance	0.50
Partial AIC	3328.24
log-likelihood ratio test	0.12 on 1 df
-log2(p) of ll-ratio test	0.46

```
Iteration 1: norm_delta = 0.43056, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 27.12197, seconds_since_start
= 0.0
Iteration 2: norm_delta = 0.04153, step_size = 0.9500, log_lik = -1635.53782, newton_decrement = 0.22899, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00238, step_size = 0.9500, log_lik = -1635.30845, newton_decrement = 0.00074, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1635.30771, newton_decrement = 0.00000, seconds_since_start =
0.1
Convergence success after 4 iterations.
```

model	lifelines.CoxPHFitter
duration col	'Months'
event col	'ID'
penalizer	0.1
l1 ratio	0.0
baseline estimation	breslow
number of observations	343
number of events observed	343
partial log-likelihood	-1635.31
time fit was run	2023-08-14 11:17:34 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
CompositeStage	0.50	1.64	0.07	0.36	0.63	1.44	1.88	0.00	7.29	<0.005	41.59

Concordance	0.63
Partial AIC	3272.62
log-likelihood ratio test	55.74 on 1 df
-log2(p) of ll-ratio test	43.46

```
Iteration 1: norm_delta = 0.13600, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 3.86282, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.01328, step_size = 0.9500, log_lik = -1659.23281, newton_decrement = 0.03364, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00074, step_size = 0.9500, log_lik = -1659.19915, newton_decrement = 0.00010, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1659.19905, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.
```

model	lifelines.CoxPHFitter
duration col	'Months'
event col	'ID'
penalizer	0.1
l1 ratio	0.0
baseline estimation	breslow
number of observations	343
number of events observed	343
partial log-likelihood	-1659.20
time fit was run	2023-08-14 11:17:34 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
LNInvolment	-0.32	0.73	0.11	-0.54	-0.09	0.58	0.91	0.00	-2.77	0.01	7.48

Concordance	0.56
Partial AIC	3320.40
log-likelihood ratio test	7.96 on 1 df
-log2(p) of ll-ratio test	7.71

```
Iteration 1: norm_delta = 0.06577, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 0.79658, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00275, step_size = 0.9500, log_lik = -1662.38897, newton_decrement = 0.00141, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00014, step_size = 0.9500, log_lik = -1662.38756, newton_decrement = 0.00000, seconds_since_start =
0.1
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1662.38756, newton_decrement = 0.00000, seconds_since_start =
0.1
Convergence success after 4 iterations.
```

model	lifelines.CoxPHFitter
duration col	'Months'
event col	'ID'
penalizer	0.1
l1 ratio	0.0
baseline estimation	breslow
number of observations	343
number of events observed	343
partial log-likelihood	-1662.39
time fit was run	2023-08-14 11:17:34 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
Comorbidity	-0.13	0.88	0.10	-0.33	0.07	0.72	1.08	0.00	-1.26	0.21	2.27

Concordance	0.53
Partial AIC	3326.78
log-likelihood ratio test	1.58 on 1 df
-log2(p) of ll-ratio test	2.26

```
Iteration 1: norm_delta = 0.02937, step_size = 0.9500, log_lik = -1663.17959, newton_decrement = 0.15086, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00064, step_size = 0.9500, log_lik = -1663.03168, newton_decrement = 0.00008, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00003, step_size = 0.9500, log_lik = -1663.03161, newton_decrement = 0.00000, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -1663.03161, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.
```

model	lifelines.CoxPHFitter
duration col	'Months'
event col	'ID'
penalizer	0.1
l1 ratio	0.0
baseline estimation	breslow
number of observations	343
number of events observed	343
partial log-likelihood	-1663.03
time fit was run	2023-08-14 11:17:34 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
FamiliyHistoryOfCancer	0.09	1.09	0.16	-0.22	0.39	0.80	1.48	0.00	0.55	0.58	0.78

Concordance	0.50
Partial AIC	3328.06
log-likelihood ratio test	0.30 on 1 df
-log2(p) of ll-ratio test	0.77

In [5]:
```
for col in data.columns:
    if col not in ['Months', 'ID']:
        n = len(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"\nColumn: {col}")
        print(f"AIC value: {aic}")
        print(f"BIC value: {bic}")
```

```
Column: DEATH
AIC value: 3328.0632186349508
BIC value: 3331.9009490821168

Column: AGE
AIC value: 3328.0632186349508
BIC value: 3331.9009490821168

Column: SEX
AIC value: 3328.0632186349508
BIC value: 3331.9009490821168

Column: CompositeStage
AIC value: 3328.0632186349508
BIC value: 3331.9009490821168

Column: LNInvolment
AIC value: 3328.0632186349508
BIC value: 3331.9009490821168

Column: Comorbidity
AIC value: 3328.0632186349508
BIC value: 3331.9009490821168

Column: FamiliyHistoryOfCancer
AIC value: 3328.0632186349508
BIC value: 3331.9009490821168
```

In [6]:
```
# Calculate p-values for each variable
p_values = []
for col in data.columns:
    if col not in ['Months', 'DEATH']:
        cph_univariate = CoxPHFitter(penalizer=0.1)
        cph_univariate.fit(data[['Months', 'DEATH', col]], duration_col='Months', event_col='DEATH', show_progress=True)
        p_values.append((col, cph_univariate.summary['p'][col]))
```

```
# Sort the p-values list in ascending order
p_values.sort(key=lambda x: x[1])

# Get the significant variable with the lowest p-value
significant_variable_pvalue = p_values[0][0]
significant_variable_pvalue_value = data[significant_variable_pvalue].iloc[0]
print(f"\nSignificant variable based on p-value: {significant_variable_pvalue}")
#print(f"Value of the significant variable: {significant_variable_pvalue_value}")
```

```
Iteration 1: norm_delta = 0.35071, step_size = 0.9500, log_lik = -943.70062, newton_decrement = 15.41644, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.01406, step_size = 0.9500, log_lik = -928.49246, newton_decrement = 0.02456, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00073, step_size = 0.9500, log_lik = -928.46793, newton_decrement = 0.00007, seconds_since_start =
0.1
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -928.46787, newton_decrement = 0.00000, seconds_since_start =
0.1
Convergence success after 4 iterations.
Iteration 1: norm_delta = 0.09094, step_size = 0.9500, log_lik = -943.70062, newton_decrement = 0.84604, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00646, step_size = 0.9500, log_lik = -942.84679, newton_decrement = 0.00411, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00033, step_size = 0.9500, log_lik = -942.84269, newton_decrement = 0.00001, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -942.84268, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.
Iteration 1: norm_delta = 0.02152, step_size = 0.9500, log_lik = -943.70062, newton_decrement = 0.04824, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00117, step_size = 0.9500, log_lik = -943.65238, newton_decrement = 0.00014, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00006, step_size = 0.9500, log_lik = -943.65224, newton_decrement = 0.00000, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -943.65224, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.
Iteration 1: norm_delta = 0.60207, step_size = 0.9500, log_lik = -943.70062, newton_decrement = 33.86694, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.06882, step_size = 0.9500, log_lik = -909.00329, newton_decrement = 0.38483, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00442, step_size = 0.9500, log_lik = -908.61639, newton_decrement = 0.00154, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -908.61485, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.
Iteration 1: norm_delta = 0.20088, step_size = 0.9500, log_lik = -943.70062, newton_decrement = 4.61862, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.02598, step_size = 0.9500, log_lik = -938.91722, newton_decrement = 0.06679, seconds_since_start =
0.0
```

```
Iteration 3: norm_delta = 0.00161, step_size = 0.9500, log_lik = -938.85015, newton_decrement = 0.00025, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -938.84989, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.
Iteration 1: norm_delta = 0.07263, step_size = 0.9500, log_lik = -943.70062, newton_decrement = 0.54710, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00311, step_size = 0.9500, log_lik = -943.15727, newton_decrement = 0.00102, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00016, step_size = 0.9500, log_lik = -943.15625, newton_decrement = 0.00000, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -943.15625, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.
Iteration 1: norm_delta = 0.02707, step_size = 0.9500, log_lik = -943.70062, newton_decrement = 0.07765, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.00205, step_size = 0.9500, log_lik = -943.62205, newton_decrement = 0.00043, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00011, step_size = 0.9500, log_lik = -943.62163, newton_decrement = 0.00000, seconds_since_start =
0.0
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -943.62163, newton_decrement = 0.00000, seconds_since_start =
0.0
Convergence success after 4 iterations.

Significant variable based on p-value: CompositeStage
```

In [7]:
```
cph_univariate = CoxPHFitter(penalizer=0.1)
cph_univariate.fit(data[['Months', 'DEATH', significant_variable_pvalue]], duration_col='Months', event_col='DEATH', show_prog
univariate_results = cph_univariate.print_summary()

# Print the univariate analysis result
print(univariate_results)
```

```
Iteration 1: norm_delta = 0.60207, step_size = 0.9500, log_lik = -943.70062, newton_decrement = 33.86694, seconds_since_start =
0.0
Iteration 2: norm_delta = 0.06882, step_size = 0.9500, log_lik = -909.00329, newton_decrement = 0.38483, seconds_since_start =
0.0
Iteration 3: norm_delta = 0.00442, step_size = 0.9500, log_lik = -908.61639, newton_decrement = 0.00154, seconds_since_start =
0.1
Iteration 4: norm_delta = 0.00000, step_size = 1.0000, log_lik = -908.61485, newton_decrement = 0.00000, seconds_since_start =
0.1
Convergence success after 4 iterations.
```

model	lifelines.CoxPHFitter
duration col	'Months'
event col	'DEATH'
penalizer	0.1
l1 ratio	0.0
baseline estimation	breslow
number of observations	343
number of events observed	176
partial log-likelihood	-908.61
time fit was run	2023-08-14 11:17:48 UTC

	coef	exp(coef)	se(coef)	coef lower 95%	coef upper 95%	exp(coef) lower 95%	exp(coef) upper 95%	cmp to	z	p	-log2(p)
CompositeStage	0.71	2.03	0.09	0.54	0.88	1.71	2.41	0.00	8.06	<0.005	50.27

Concordance	0.68
Partial AIC	1819.23
log-likelihood ratio test	70.17 on 1 df
-log2(p) of ll-ratio test	54.03

None

In [9]:
```
n = len(data)
llf = cph_univariate.log_likelihood_
k = cph_univariate.params_.shape[0]
aic = -2 * llf + 2 * k
bic = -2 * llf + k * np.log(n)
```

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

```
AIC value: 1819.2296914692981
BIC value: 1823.0674219164641
```

In [ ]: