

# *Lung cancer*

GEO DATASET ANALYSIS

---

# Table of Contents

|   |                              |   |            |
|---|------------------------------|---|------------|
| 1 | Introduction                 | 5 | Histology  |
| 2 | Project information          | 6 | Conclusion |
| 3 | Data acquisition             | 7 | Contact    |
| 4 | Cell size:<br>NSCLC and SCLC |   |            |

---

## Introduction

# Lung cancer

폐암은 폐에 생기는 악성 종양으로, 발생 위치나 치료 방식에 따라 소세포폐암(SCLC)과 비소세포폐암(NSCLC)으로 구분된다.

암종별(SCLC vs NSCLC) 및 세부 병리학적 조직학(Histology) 분류에 따라 생존 곡선과 평균 생존 시간의 차이가 실제 통계적으로 유의미한지 분석하고자 하였다.

이 분석의 가설은 다음과 같다.

귀무가설: 폐암의 Subtype과 상관없이 모든 그룹의 생존 기간 분포가 동일하다

대립가설: 적어도 하나의 Subtype 그룹은 다른 그룹과 생존 기간 분포가 다르다

## Introduction

# Project information

Data resource: NCBI GEO (GSE30219)

### Libraries

- lifelines
- Biopython
- Pandas
- Seaborn
- GEOparse
- Scipy.stats
- scikit\_posthocs

### Statics

- Mann-Whitney U test
- Kruskal-Wallis test
- Dunn test

# Flowchart



## Data acquisition

1

Data acquisition

GEOparse를 이용해 데이터 입수

GSE30219

2

Data preprocessing

필요한 칼럼 정제

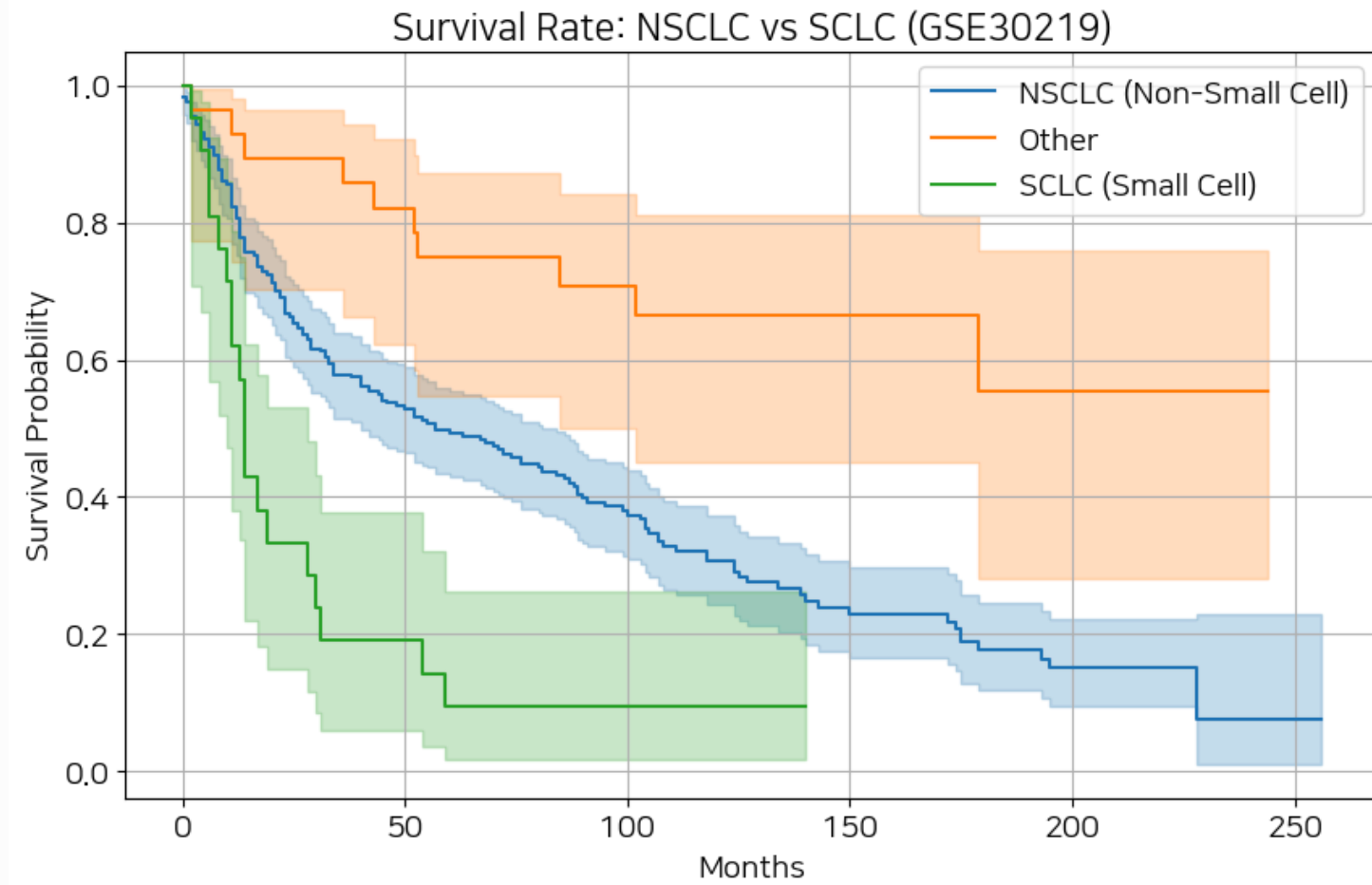
Histology: 암종

Month: 생존 기간

Status: 생존 상태

# Cell size: NSCLC and SCLC

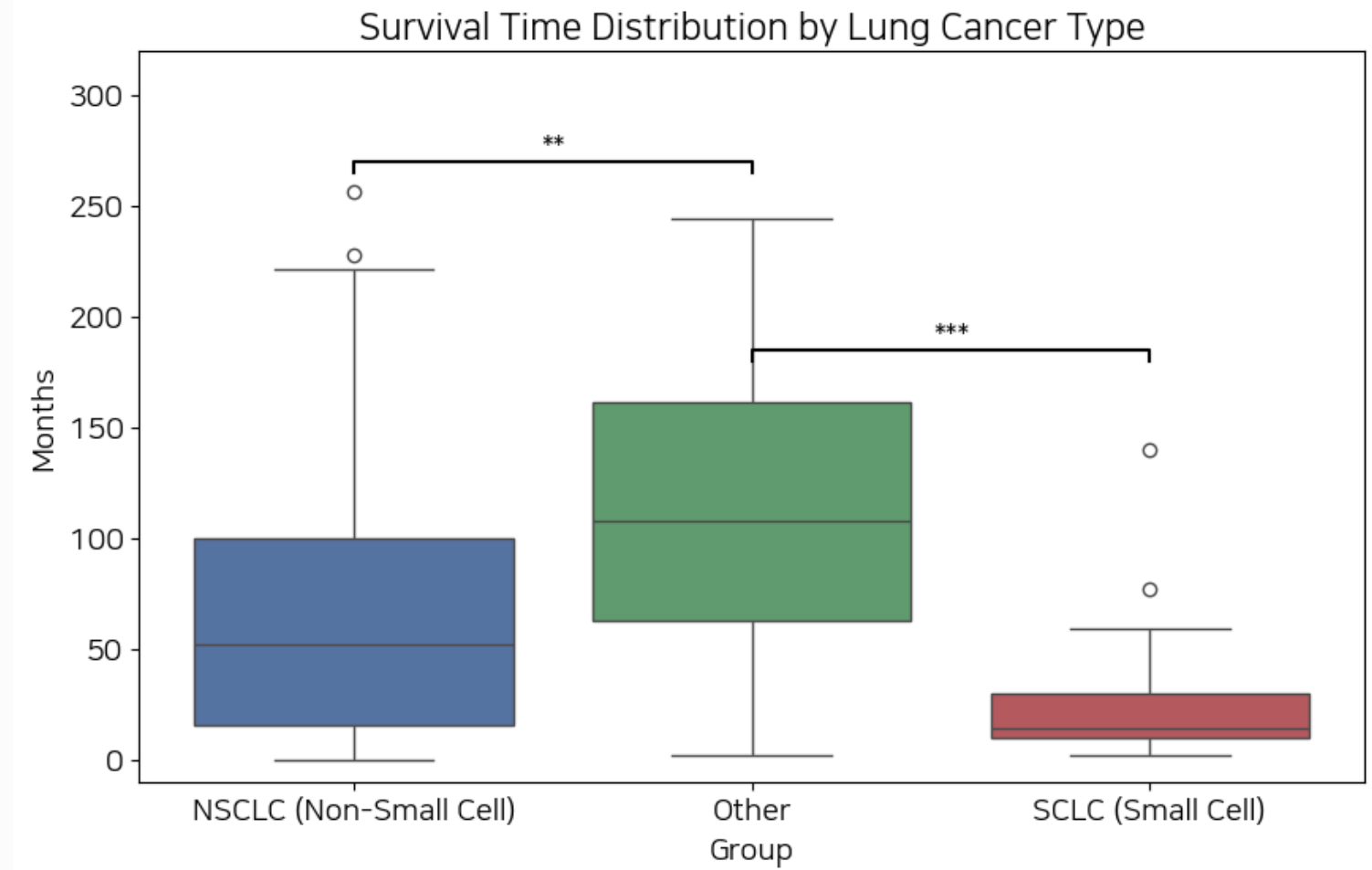
## Lifelines & Boxplot



Lifelines

## 비소세포성 폐암과 소세포성 폐암간 비교

50개월 시점에서 소세포성 폐암의 생존율이 20% 미만으로 떨어짐을 확인



Boxplot

## Cell size: NSCLC and SCLC

### Statistics

귀무가설: Subtype과 상관 없이 모든 그룹의 생존 기간 분포가 동일하다.

대립가설: 적어도 하나의 Subtype 그룹은 다른 그룹과 생존 기간 분포가 다르다.

#### Kruskal-Wallis Test

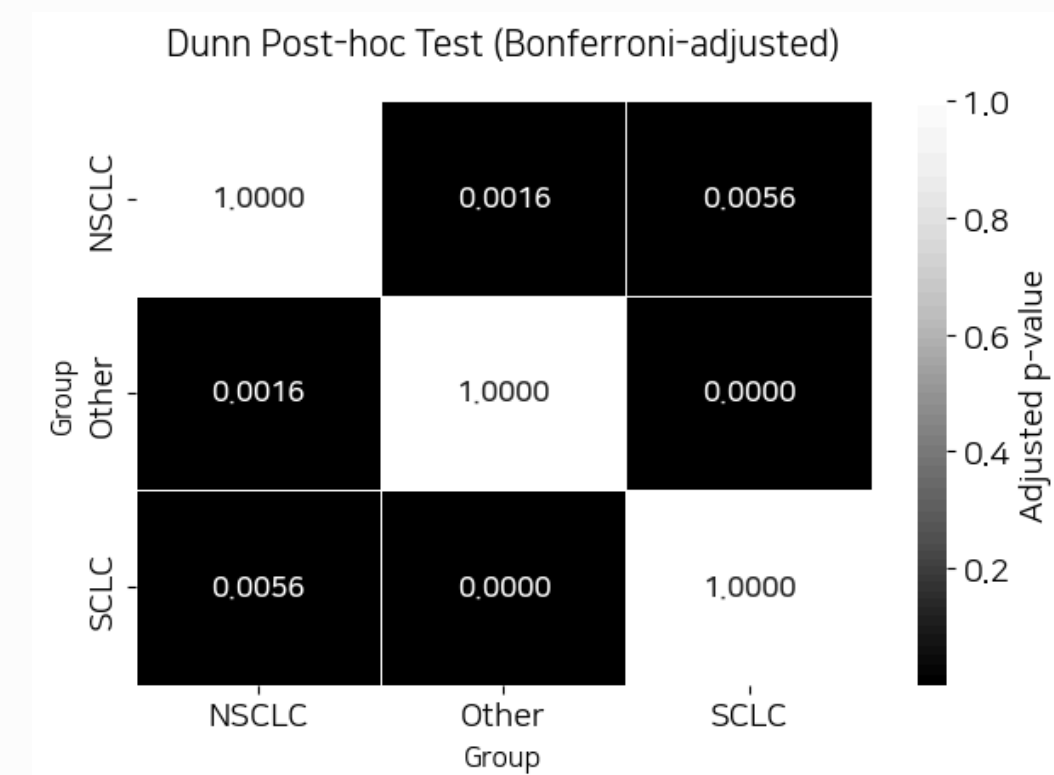
| Kruskal-Wallis | P-value    |
|----------------|------------|
| 23.8333        | $p < 0.05$ |

적어도 하나의 Subtype 그룹은 다른 그룹과 생존 기간 분포가 다름을 확인, 후속 분석으로 Dunn's test를 진행함.

### Kruskal-Wallis Test & Dunn's test

샤피로-윌크 검정과 레빈 검정을 통해 정규성, 등분산성을 위반함을 확인, 따라서 비모수검정으로 진행

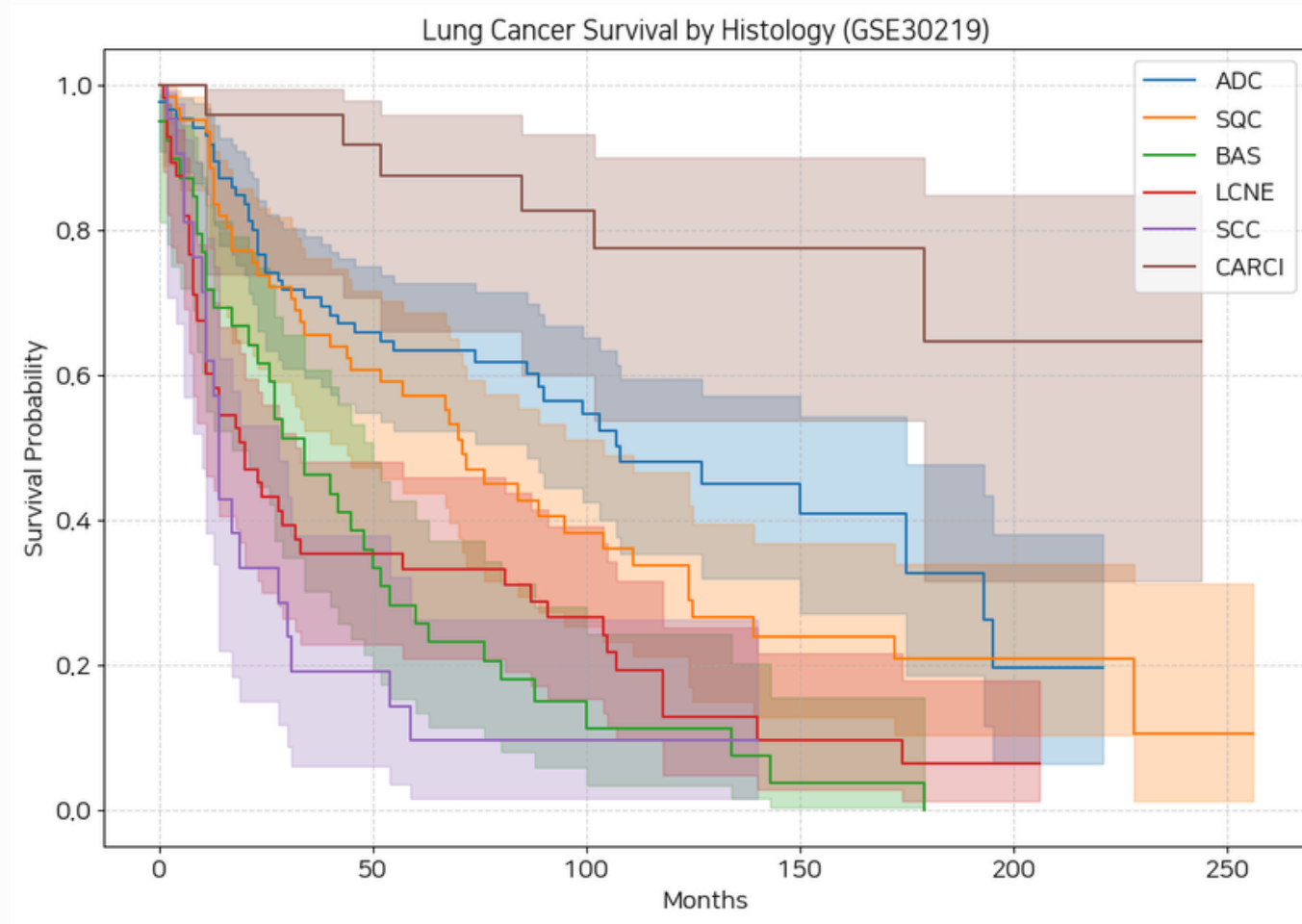
#### Dunn's test



모든 비교 쌍에서  $p\text{-value} < 0.05$ 이므로, 각 그룹 간의 생존 기간 차이가 통계적으로 매우 유의미함을 확인

# Histology

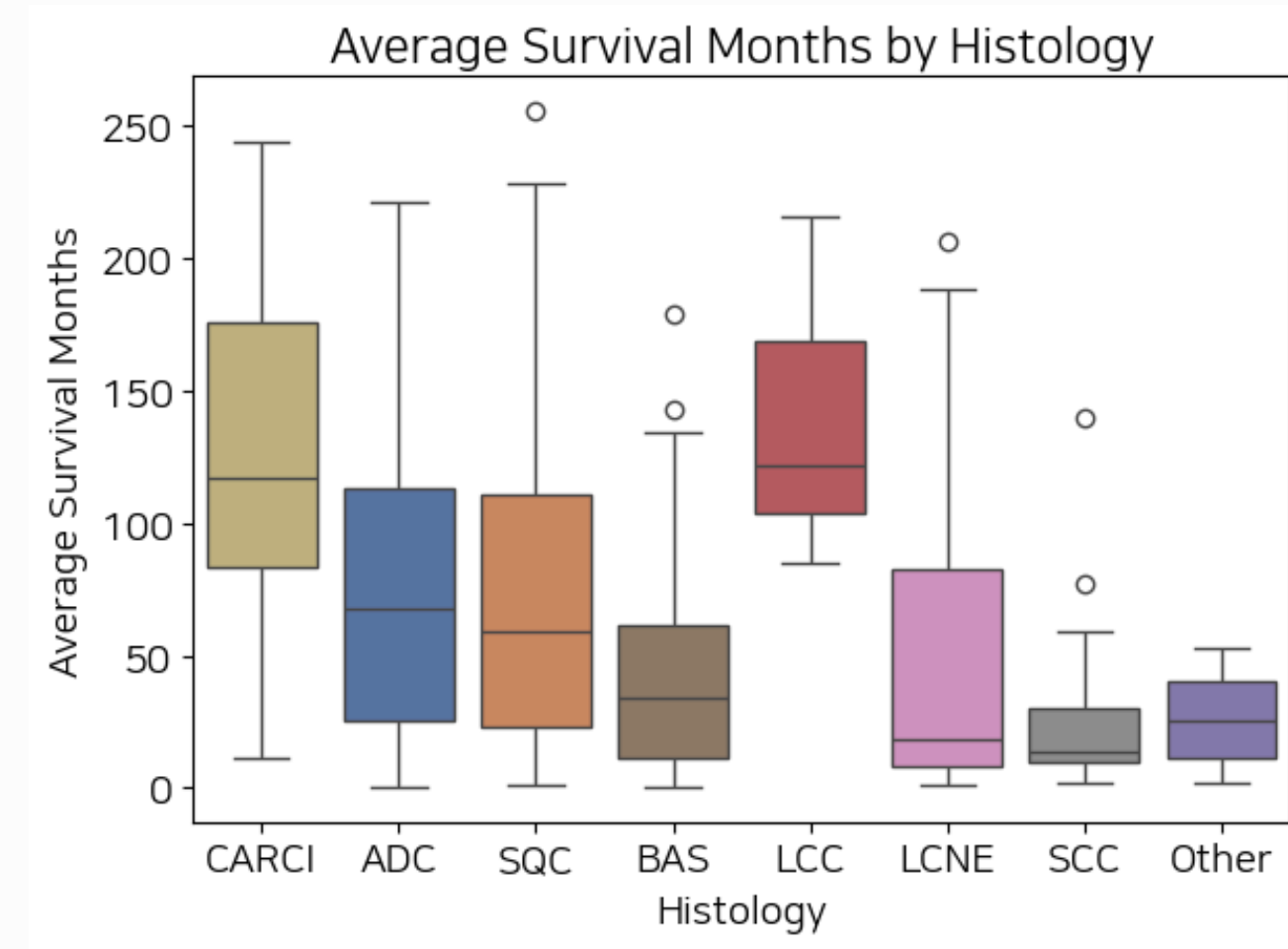
## Lifelines & Boxplot



Lifelines

## Histology별 비교

SCC(소세포암)의 중앙값이 약 15~20개월로 가장 낮았음.



Boxplot

## Conclusion

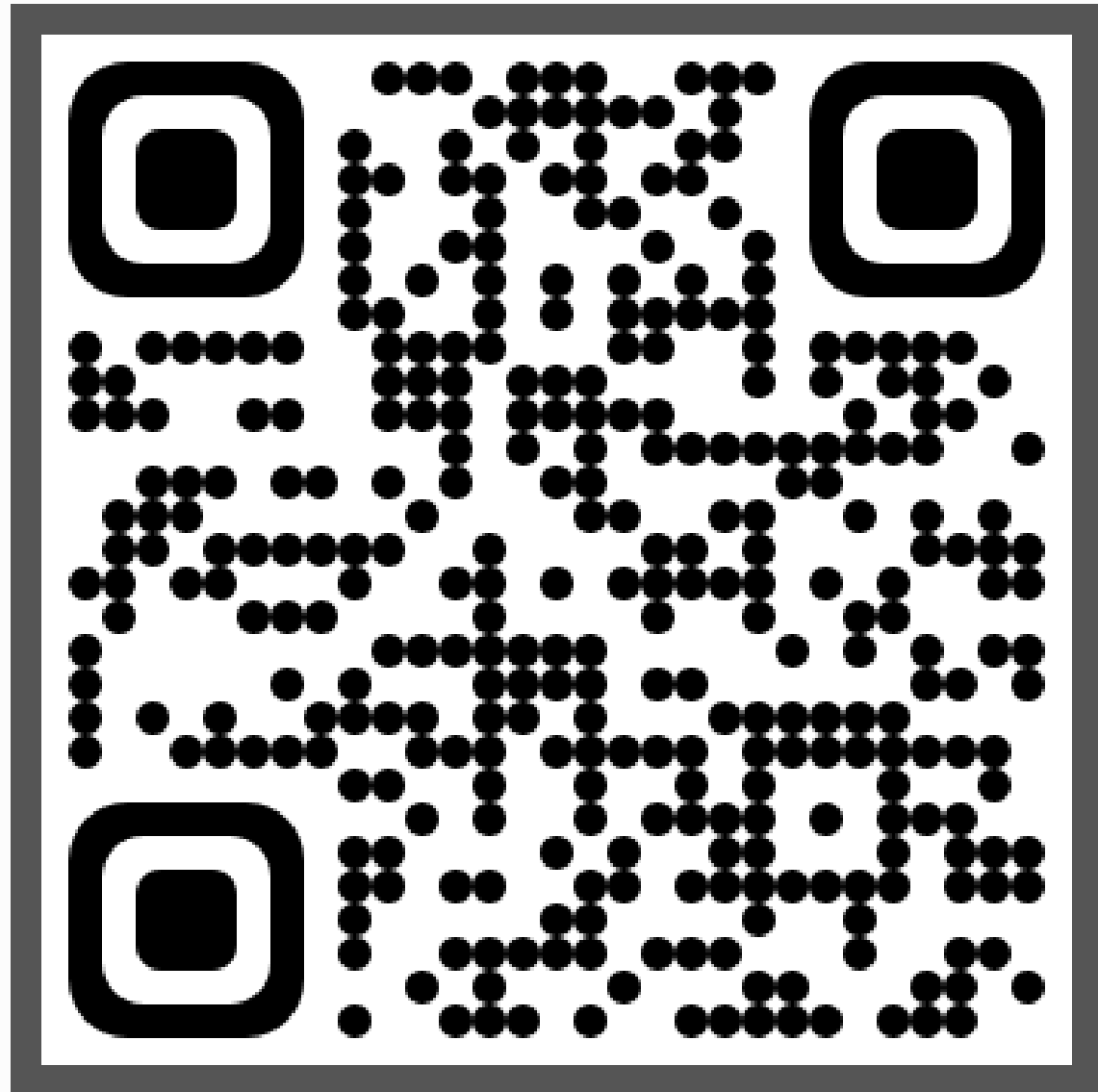
# Conclusion

50개월 시점에서 SCLC의 생존율은 20% 미만으로 떨어졌으며, 평균 생존 기간 역시 중앙값이 약 15~20개월 이내로 암종들 중에서 가장 짧았다.

- 1.Kruskal-Wallis Test를 통해 적어도 하나의 subtype 그룹은 다른 그룹과 생존 기간의 분포가 다름을 통계적으로 확인
- 2.Dunn's test를 통해 subtype간 생존 기간 분포 차이가 통계적으로 유의함을 확인

Histology별 분석을 진행했을 때도 CARCI(유암종) 그룹은 평균 생존 월수 중앙값이 약 120개월에 달하며 예후가 가장 좋은 반면, SCC(소세포암)와 LCNE(대세포 신경내분비암)는 중앙값이 20개월 미만에 머물렀음.

## Contact



### Mail

pokemonms@naver.com  
blackholekun@gmail.com

### Cellular

+82-10-5027-0328

### Github

<https://github.com/koreanraichu>

### Blog

<https://koreanraichu.tistory.com/>