

Analysis of stage III stomach cancer using the restricted mean survival time

Bitna Kim^a, Minjung Lee^{1,a}

^aDepartment of Statistics, Kangwon National University

Abstract

The purpose of this study is to compare the effects of treatment on stage III stomach cancer data obtained from the SEER program of the National Cancer Institute and to identify the significant risk factors for the survival rates of stage III stomach cancer. Since the proportional hazards assumption was violated for treatment, we used the restricted mean survival time as an alternative to the proportional hazards model. The restricted mean survival time was estimated using pseudo-observations, and the effects of treatment were compared using a test statistic based on the estimated restricted mean survival times. We conducted the regression analysis using a generalized linear model to investigate the significant predictors for the restricted mean survival time of patients with stage III stomach cancer. We found that there was a significant difference between the restricted mean survival times of treatment groups. Age at diagnosis, race, substage, grade, tumor size, surgery, and treatment were significant predictors for the restricted mean survival time of patients with stage III stomach cancer. Surgery was the most significant predictor for increasing the restricted mean survival time of patients with stage III stomach cancer.

Keywords: pseudo-observations, restricted mean survival time, stomach cancer, survival analysis

1. 서론

위암(stomach cancer)은 위에서 발생하는 악성 종양으로 국제질병분류(International Classification of Diseases for Oncology) C16으로 나타내어진다. 보건복지부 중앙암등록본부의 2017년 통계에 따르면 위암은 전체 암 발생자 232,255명 중 29,685명으로 암 발생율 1위를 차지하였다 (Korea Central Cancer Registry, 2019). 위암의 발병요인으로는 가족력, 헬리코박터 파일로리균(*Helicobacter pylori*) 감염, 위 관련 기존 질병과 식습관 등이 있으며 최근 한국인들이 육류와 가공식품을 즐겨 먹는 식습관이 위암 발생율 증가에 크게 영향을 미치고 있다. 위암은 증상이 없는 대표적인 암으로 과거에는 진단이 늦어져 치명적이었으나 최근에는 국내 6대암(위암, 폐암, 간암, 대장암, 유방암, 자궁경부암)으로 지정되는 등 국가적인 지원과 진단 및 치료기술의 발달로 위암의 생존율은 증가하는 추세이다. 그러나 기수가 높아질수록 위암의 5년 상대생존율은 급격히 감소하므로 (National Cancer Information Center, 2019) 위암 생존율에 관한 꾸준한 연구가 필요하다.

This work was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (NRF-2018R1D1A1B07041070).

This paper is a condensed form of the first author's master thesis from the Kangwon National University, Chuncheon, Korea.

¹ Corresponding author: Department of Statistics, Kangwon National University, 1, Kangwondaehak-gil, Chuncheon-si, Gangwon-do 24341, Korea. E-mail: mlee@kangwon.ac.kr

위암 환자의 생존율에 영향을 미치는 요인으로 위암의 진행 상황이 있으며, 위암의 진행은 병기(stage)와 분화도(grade)로 설명된다. 위암의 병기는 크게 조기 위암과 진행성 위암으로 나뉘며 1기에서 4기까지 나누어져 있다. 이는 종양의 장벽 침윤정도, 림프절 전이 개수, 원격전이 여부에 의해 분류되는 TNM 병기 분류에 의해 세분화되어 있다. 위암 3기의 경우 IIIA와 IIIB로 세분화되어지며 기수가 높아질수록 침범깊이가 깊고 림프절 전이 개수가 많으며, 위암 4기는 원격전이까지 발생한 상태로 기수가 높아질수록 치명적인 요인이 많아진다. 암의 분화도는 현미경으로 관찰한 종양세포의 모양에 따른 분류로 고분화 단계(well-differentiated cancer)의 암은 낮은 등급의 분화도에 속하며, 미분화 단계(undifferentiated cancer)의 암은 높은 등급의 분화도에 속한다. 미분화 단계의 암은 고분화 단계의 암보다 빠른 속도로 분열하고 전이가 쉽게 일어난다 (Ruhl 등, 2019).

암의 진행 상황 외에도 나이, 성별, 인종, 기혼여부, 진단연도 등과 같은 환자의 정보와 종양의 크기와 암의 치료방법 또한 위암 생존율에 영향을 미치는 요인으로 고려될 수 있다. 위암의 치료방법으로는 수술과 항암치료인 화학요법과 방사선치료 등이 있다. 위암에서 수술이 가장 직접적이고 먼저 고려되는 치료방법이기는 하나 항암치료가 수술 전후 환자의 상태에 영향을 주고 수술보다 지속적으로 환자의 상태에 영향을 주는 치료법이므로 수술만큼 중요하다.

본 논문에서는 미국 국립암연구소(National Cancer Institute)의 SEER(Surveillance, Epidemiology, and End Results) 프로그램에서 제공하는 위암 3기 자료를 분석하여 항암치료의 효과를 비교하고 위암 3기 환자의 생존에 유의한 영향을 미치는 요인들을 파악하고자 한다. 항암치료에 따른 위암 생존율을 비교하고 위암 생존율에 유의한 영향을 미치는 공변량들(covariates)의 효과를 추정할 때, 생존분석에서 가장 많이 사용되는 분석 방법은 로그-순위 검정(log-rank test)과 Cox 비례위험모형(proportional hazards model) (Cox, 1972)이다. 로그-순위 검정과 Cox 비례위험모형은 위험비가 시간에 의존하지 않고 일정하다는 비례위험 가정이 성립할 때 사용할 수 있는 분석법이며, 비례위험 가정이 성립하지 않을 때는 그 결과를 신뢰할 수 없다. 이에 대한 대안으로 제한된 평균 생존시간(restricted mean survival time)을 사용할 수 있다. 제한된 평균 생존시간은 사건발생시간과 미리 정해진 관심의 시점의 최소값의 기댓값으로 정의되며 (Irwin, 1949), 연구시작시점 0에서부터 미리 정해진 관심의 시점까지의 생존함수 곡선 아래 면적으로 표현되어 해석이 직관적이며 임상학적으로 유의미한 해석을 할 수 있다는 장점이 있다 (Chen과 Tsiatis, 2001; Andersen 등, 2004; Royston과 Parmar, 2011; Zhao 등, 2012; Tian 등, 2014). 제한된 평균 생존시간은 비례위험 가정이 성립하지 않을 때 사용할 수 있을 뿐만 아니라 비례위험 가정이 성립할 때에도 로그-순위 검정, 비례위험모형과 함께 이차적인 측도로 활용가능하다. 본 논문에서는 위암 3기 자료에 포함된 공변량이 비례위험 가정을 만족하지 않을 경우 대안으로 제한된 평균 생존시간을 이용하여 항암치료의 효과를 비교하고 위암 3기 환자의 평균 생존시간에 유의한 영향을 미치는 요인들을 파악하고자 한다.

본 논문의 구성은 다음과 같다. 2절에서는 제한된 평균 생존시간의 정의와 집단 간 제한된 평균 생존시간을 비교하기 위한 검정통계량을 소개하고, 제한된 평균 생존시간의 회귀모형과 회귀모수 추정에 대해 설명한다. 3절에서는 비례위험 가정이 성립하지 않는 위암 3기 환자 자료를 제한된 평균 생존시간을 이용하여 분석한 결과를 보여준다. 4절에서는 결론을 제시하며 마무리한다.

2. 연구 방법

2.1. 제한된 평균 생존시간의 정의

연구 시작 후 관심 사건이 발생하기까지 걸린 시간을 T , 우중도 절단시간(right censoring time)을 C , 관측시간을 $X = \min(T, C)$ 라고 하자. p 개의 공변량들의 벡터를 \mathbf{Z} 라고 하면, 관측된 자료는 $(X_i, \delta_i, \mathbf{Z}_i)$ ($i = 1, \dots, n$)

로 나타낼 수 있다. 여기서 $\delta = I(T \leq C)$ 는 사건발생 여부를 나타내는 지시자이며, $I(\cdot)$ 는 지시함수(indicator function)이다. 공변량 \mathbf{Z} 가 주어져 있을 때, 사건발생시간 T 와 우중도 절단시간 C 는 독립이라고 가정한다.

$P(X \geq \tau) > 0$ 인 시점 τ 를 연구자가 미리 정한 관심의 시점이라고 하자. 제한된 평균 생존시간 $\mu(\tau)$ 는 사건 발생시간 T 와 미리 정한 관심의 시점 τ 의 최소값의 기댓값으로 정의된다 (Irwin, 1949). 즉, $\mu(\tau) = E\{\min(T, \tau)\}$ 이며, 다음과 같이 0부터 τ 까지의 생존함수 $S(t) = P(T > t)$ 곡선의 아래 면적으로 표현될 수 있다.

$$\mu(\tau) = \int_0^{\tau} S(t)dt.$$

공변량 \mathbf{Z} 가 주어져 있을 때, 제한된 평균 생존시간은 $\mu(\tau; \mathbf{Z}) = E\{\min(T, \tau)|\mathbf{Z}\}$ 로 정의된다.

2.2. 의사-관측들을 이용한 제한된 평균 생존시간 추정

의사-관측들(pseudo-observations)을 이용하여 제한된 평균 생존시간 $\mu(\tau)$ 를 추정할 수 있다. 모든 개체들에 대해 leave-one-out 방법을 통해 $\mu(\tau)$ 의 의사-관측치들을 계산하고, 의사-관측치들의 평균으로 $\mu(\tau)$ 를 추정한다. i 번째 의사-관측 $\hat{\theta}_i(\tau)$ 는 다음과 같이 계산된다 (Andersen 등, 2004).

$$\hat{\theta}_i(\tau) = n \int_0^{\tau} \hat{S}(t)dt - (n-1) \int_0^{\tau} \hat{S}^{(-i)}(t)dt, \quad (2.1)$$

여기서 $\hat{S}(t)$ 는 Kaplan-Meier 추정량 (Kaplan과 Meier, 1958)이며, $\hat{S}^{(-i)}(t)$ 는 i 번째 관측값을 제외한 후 구한 Kaplan-Meier 추정량이다. 식 (2.1)의 의사-관측들을 이용하여 제한된 평균 생존시간 추정량 $\hat{\mu}(\tau)$ 와 그 분산추정량 $\hat{\sigma}^2(\tau)$ 는 다음과 같이 구할 수 있다 (Nemes 등, 2020).

$$\hat{\mu}(\tau) = \frac{1}{n} \sum_{i=1}^n \hat{\theta}_i(\tau), \quad \hat{\sigma}^2(\tau) = \frac{1}{n} \frac{1}{n-1} \sum_{i=1}^n \{\hat{\theta}_i(\tau) - \hat{\mu}(\tau)\}^2. \quad (2.2)$$

2.3. 집단 간 제한된 평균 생존시간 비교

K 개 집단 간의 제한된 평균 생존시간을 비교하기 위한 귀무가설과 대립가설은 다음과 같다.

$$H_0 : \mu_1(\tau) = \mu_2(\tau) = \dots = \mu_K(\tau)$$

$$H_1 : \text{적어도 하나의 } \mu_k(\tau) \text{가 다르다.}$$

위 귀무가설을 검정하기 위한 검정통계량은 다음과 같다 (Guo와 Liang, 2019).

$$\chi^2 = \hat{\boldsymbol{\mu}}(\tau)^T D^T (D \hat{\boldsymbol{\Sigma}} D^T)^{-1} D \hat{\boldsymbol{\mu}}(\tau) \sim \chi^2(df), \quad (2.3)$$

여기서 $\hat{\boldsymbol{\mu}}(\tau) = (\hat{\mu}_1(\tau), \hat{\mu}_2(\tau), \dots, \hat{\mu}_K(\tau))^T$ 는 K 개 집단의 제한된 평균 생존시간 추정량들의 벡터이며, D 는 j 번째 행이 $e_j - e_{j+1}$ 인 $(K-1) \times K$ 행렬이며, e_j 는 j 번째 원소만 1이고 나머지는 0인 $K \times 1$ 벡터이다. $\hat{\boldsymbol{\Sigma}}$ 는 $\hat{\boldsymbol{\mu}}(\tau)$ 의 공분산 행렬로 $K \times K$ 대각 행렬이며 j 번째 대각 원소가 $\hat{\mu}_j(\tau)$ 의 분산추정량인 $\hat{\sigma}_j^2(\tau)$ 이다. 식 (2.3)의 검정통계량은 귀무가설 하에서 χ^2 분포를 따르며 자유도 df 는 $K-1$ 이다. 유의수준 α 에서 $\chi^2 > \chi_{\alpha}^2(K-1)$ 이면 귀무가설 H_0 를 기각하고 K 개 집단 간의 제한된 평균 생존시간에 유의한 차이가 있다고 할 수 있다.

2.4. 제한된 평균 생존시간의 회귀모형

제한된 평균 생존시간 $\mu(\tau; \mathbf{Z})$ 에 영향을 미치는 공변량 \mathbf{Z} 의 효과를 추정하기 위해, Andersen 등 (2004)은 다음과 같이 일반화 선형모형(generalized linear model)을 통해 제한된 평균 생존시간 $\mu(\tau; \mathbf{Z})$ 에 공변량 \mathbf{Z} 의 효과를

직접 모형화하는 것을 제안하였다.

$$g\{\mu(\tau; \mathbf{Z}_i)\} = \boldsymbol{\beta}^T \mathbf{W}_i \quad (2.4)$$

여기서 $g(\cdot)$ 은 연결함수(link function)이며, $\boldsymbol{\beta}$ 는 회귀모수들의 $(p+1) \times 1$ 벡터이며, $\mathbf{W}_i = (1, \mathbf{Z}_i^T)^T$ 이다. \mathbf{W}_i 의 첫 번째 원소는 절편에 해당하며, 항등함수(identity function), 로그함수(log function)가 연결함수 $g(\cdot)$ 로 사용될 수 있다 (Andersen과 Perme, 2010; Guo와 Liang, 2019). Andersen 등 (2004)은 식 (2.1)의 의사-관측들을 이용하여 다음과 같이 일반화 추정 방정식(generalized estimating equation)의 해로서 식 (2.4)의 회귀모수 $\boldsymbol{\beta}$ 를 추정하였다.

$$U(\boldsymbol{\beta}) = \sum_{i=1}^n U_i(\boldsymbol{\beta}) = \sum_{i=1}^n \left(\frac{\partial}{\partial \boldsymbol{\beta}} g^{-1}(\boldsymbol{\beta}^T \mathbf{W}_i) \right)^T V_i^{-1} (\hat{\theta}_i(\tau) - g^{-1}(\boldsymbol{\beta}^T \mathbf{W}_i)) = 0,$$

여기서 V_i 는 $\hat{\theta}_i(\tau)$ 의 가분산(working variance)이며, $\mu(\tau; \mathbf{Z}_i)$ 가 다변량인 경우에는 V_i 는 $\hat{\theta}_i(\tau)$ 의 가공분산 행렬(working covariance matrix)이다 (Liang 과 Zeger, 1986; Zeger 와 Liang, 1986). Andersen 등 (2004)은 간단하게 V_i 를 $V_i = 1$ 또는 항등행렬(identity matrix) I 로 설정하였다. 회귀모수 추정량 $\hat{\boldsymbol{\beta}}$ 의 공분산 행렬 $\hat{\Sigma}$ 는 다음과 같이 추정된다.

$$\hat{\Sigma} = \mathbf{I}(\hat{\boldsymbol{\beta}})^{-1} \widehat{\text{Var}}\{U(\hat{\boldsymbol{\beta}})\} \mathbf{I}(\hat{\boldsymbol{\beta}})^{-1},$$

여기서 $\mathbf{I}(\boldsymbol{\beta}) = \sum_{i=1}^n (\partial g^{-1}(\boldsymbol{\beta}^T \mathbf{W}_i) / \partial \boldsymbol{\beta})^T V_i^{-1} (\partial g^{-1}(\boldsymbol{\beta}^T \mathbf{W}_i) / \partial \boldsymbol{\beta})$, $\widehat{\text{Var}}\{U(\hat{\boldsymbol{\beta}})\} = \sum_{i=1}^n U_i(\hat{\boldsymbol{\beta}}) U_i(\hat{\boldsymbol{\beta}})^T$ 이다.

3. 위암 3기 자료 분석

3.1. 자료 설명

본 논문에서 분석한 자료는 미국 국립암연구소의 SEER 프로그램에서 제공하는 2004년에서 2015년까지 위암 3기 진단을 받은 19세에서 94세까지의 성인 환자 3,023명의 자료이다 (SEER, 2019). 최대 관측시간은 154개월이며, 3,023명의 위암 3기 환자 중 1,966명이 위암으로 인하여 사망하였고 나머지 1,057명은 중도절단되었다. Table 1과 Table 2는 분석에 사용된 공변량들의 기술통계를 보여준다.

3.2. 로그-순위 검정과 Cox 비례위험모형 적합 결과

항암치료로 화학요법과 방사선치료가 사용되었으며 항암치료에 따라 생존율에 유의한 차이가 있는지 알아 보기 위해 로그-순위 검정을 시행한 결과, 유의수준 5%에서 항암치료법에 따라 생존율에서 유의한 차이가 있는 것으로 나타났다(p -value < 0.001). Figure 1은 항암치료법에 따른 Kaplan-Meier 생존곡선을 보여준다. 화학요법만 받은 환자들의 생존곡선과 방사선치료만 받은 환자들의 생존곡선이 57개월에서 교차하고 방사선치료만 받은 환자들의 생존곡선과 화학요법과 방사선치료 모두 받은 환자들의 생존곡선이 125개월에서 교차하므로 항암치료법이 비례위험 가정을 만족하지 않음을 예상할 수 있다.

성별, 인종, 기혼여부, 진단연도, 진단연령, 세분화병기, 분화도, 종양의 크기, 수술여부, 항암치료가 위암 사망 위험률에 미치는 영향을 파악하기 위해 Cox 비례위험모형을 적합하였다. 종양의 크기는 오른쪽으로 꼬리가 긴 분포를 가지고 있어 정규화를 위해 로그 변환하여 사용하였다. 비례위험모형 적합 결과, 유의수준 5%에서 진단연령, 인종, 세분화병기, 분화도, 종양의 크기, 수술여부, 항암치료는 위암 3기 환자의 생존율에 유의한 영향을 미치는 요인들이었으나, 진단연도, 성별, 기혼여부는 유의한 요인이 아니었다. Table 3은 유의수준 5%에서 유의한 변수만을 모형에 포함시키고 비례위험모형을 적합한 결과이다. 진단시 연령이 높을수록

Table 1: Data characteristics

Categorical variables		<i>n</i>	%
Sex	Female	867	28.68
	Male	2156	71.32
Race	White	2033	67.25
	Black	358	11.84
	Others	632	20.91
Marital status	Married	2500	82.70
	Single	523	17.30
Substage	IIIA	2398	79.33
	IIIB	625	20.67
Grade	Low	675	22.33
	High	2348	77.67
Surgery	Not performed	226	7.48
	Performed	2797	92.52
Treatment	No therapy	716	23.69
	Chemotherapy	620	20.51
	Radiotherapy	118	3.90
	Chemotherapy + Radiotherapy	1569	51.90

Table 2: The mean and standard deviation of age and tumor size

Continuous variables	Mean	Standard deviation
Age at diagnosis	63.44	13.12
Tumor size	57.72	36.55

종양의 크기가 클수록 세분화병기와 분화도가 높은 단계일수록 위암으로 사망할 위험률이 높아지며, 수술과 항암치료를 받은 경우 위암으로 사망할 위험률이 낮아짐을 알 수 있다.

3.3. 비례위험 가정 검토

비례위험모형에 포함된 공변량들에 대해, Schoenfeld 잔차를 기반한 χ^2 검정 (Grambsch와 Therneau, 1994) 과 Schoenfeld 잔차 (Schoenfeld, 1982) 도표, $\log[-\log\{\hat{S}(t)\}]$ 곡선을 이용하여 비례위험 가정을 검토하였다. Table 3의 Schoenfeld 잔차를 기반한 χ^2 검정 결과, 유의수준 5%에서 진단연령, 수술여부, 항암치료가 비례위험 가정을 만족하지 않는 것으로 나타났다. 이 변수들에 대해 Schoenfeld 잔차 도표를 그려 시각적으로 비례위험 가정을 검토하였다. Schoenfeld 잔차 도표에서 lowess 곡선이 어떠한 추세를 나타내지 않고 0에서 수평한 형태를 보이면 해당 공변량은 비례위험 가정을 만족한다고 할 수 있다. Figure 2에서 수술여부와 진단연령은 비례위험 가정을 만족한다고 할 수 있으나, 항암치료의 Chemotherapy, Chemotherapy+Radiotherapy의 잔차 도표에서 lowess 곡선이 증가하다가 감소하는 추세가 있으므로 $\log[-\log\{\hat{S}(t)\}]$ 곡선을 이용하여 항암치료 변수의 비례위험 가정을 한 번 더 검토하였다. Figure 2에서 항암치료 변수의 $\log[-\log\{\hat{S}(t)\}]$ 곡선이 서로 교차하는 부분이 있으므로 항암치료 변수는 비례위험 가정을 만족하지 않음을 알 수 있다.

3.4. 제한된 평균 생존시간을 이용한 항암치료의 효과 비교

항암치료 변수가 비례위험 가정을 만족하지 않으므로 대안으로 제한된 평균 생존시간을 사용하였다. 제한된 평균 생존시간을 추정하기 위해 τ 를 연구 초반인 20개월, 화학요법만 받은 환자와 방사선치료만 받은 환자의

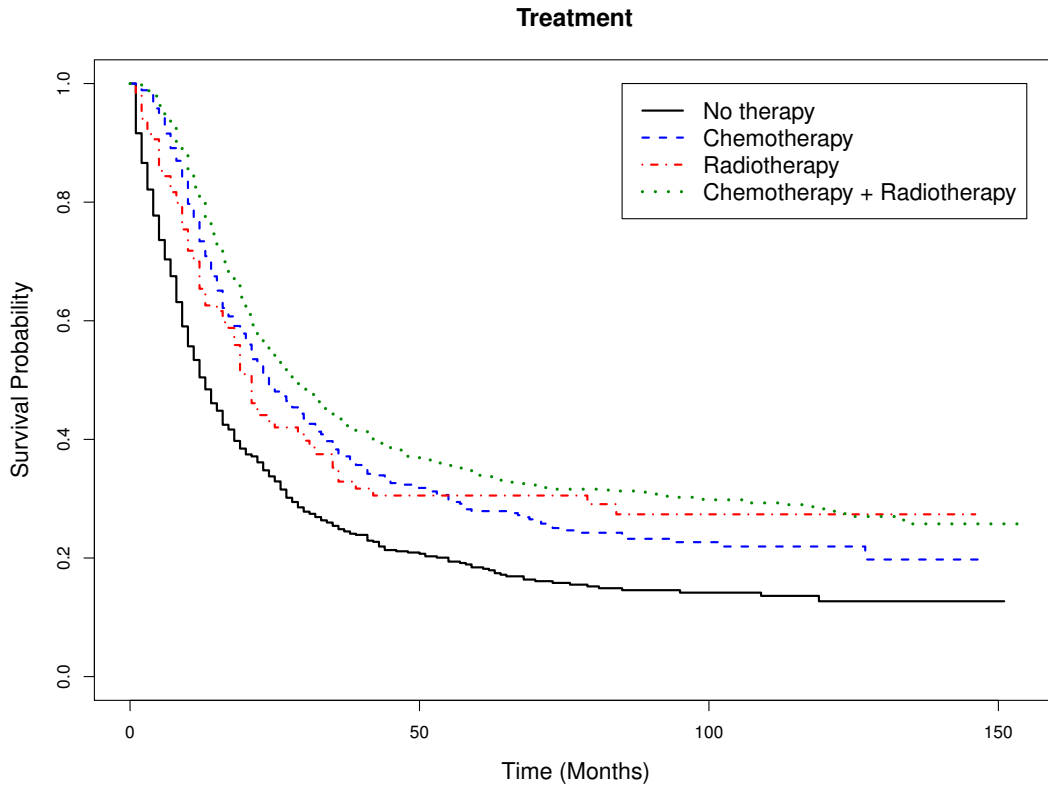


Figure 1: Estimated survival probabilities for treatment.

생존곡선이 교차하는 시점인 57개월, 연구 후반인 100개월, 위암 사망이 발생한 시점들 중 최대 시점인 134개월로 정하였다. Table 4는 $\tau = 20, 57, 100, 134$ 개월일 때, 항암치료법에 따른 위암 3기 환자의 평균 생존시간 추정값과 항암치료법에 따른 평균 생존시간을 비교하기 위한 검정통계량 χ^2 값과 p -value를 보여준다. 여기서 No therapy는 항암치료를 받지 않음(즉, 화학요법과 방사선치료 모두 받지 않음)을 의미한다. $\tau = 134$ 개월일 때, 아무런 치료도 받지 않은 위암 3기 환자의 평균 생존시간은 약 33.1개월, 화학요법만 받은 환자의 평균 생존시간은 48.6개월, 방사선치료만 받은 환자의 평균 생존시간은 50개월, 화학요법과 방사선치료 모두 받은 환자의 평균 생존시간은 56.6개월이었다. 선택한 모든 τ 에 대해, 유의수준 5%에서 항암치료법에 따라 위암 3기 환자의 평균 생존시간에는 통계적으로 유의한 차이가 있음을 확인하였다. 또한 화학요법과 방사선치료 모두 받은 환자의 평균 생존시간이 가장 길었음을 알 수 있었다.

3.5. 제한된 평균 생존시간의 회귀모형 적합결과

위암 3기 환자의 평균 생존시간에 유의한 영향을 미치는 공변량들을 파악하기 위해, 식 (2.4)에서 로그 함수를 연결함수로 사용하여 회귀모형을 적합하였다. 로그함수를 연결함수로 사용한 이유는 제한된 평균 생존시간 추정값은 음이 아닌(nonnegative) 값을 가지는데 항등 연결함수(identity link function)를 사용하면 추정값이 그 범위를 벗어날 수도 있기 때문이다. 또한 비례위험 가정이 성립하지 않을 때 또 다른 대안으로 가속화 실패 시간 모형(accelerated failure time model)을 사용할 수 있으며, 가속화 실패 시간 모형은 로그화된 생존시간에

Table 3: Regression parameter estimates, standard errors, and p -values for death from stomach cancer under the proportional hazards model, and test of the proportional hazards (PH) assumption

Covariates		Proportional hazards model			PH assumption test		
		$\hat{\beta}$	se($\hat{\beta}$)	p -value	χ^2	df	p -value
Age at diagnosis		0.013	0.002	< 0.001	40.795	1	< 0.001
Race	White	0	–	–	0	–	–
	Black	0.118	0.070	< 0.001	0.005	1	0.944
	Others	-0.347	0.060	< 0.001	0.225	1	0.635
Substage	IIIA	0	–	–	0	–	–
	IIIB	0.270	0.055	< 0.001	0.070	1	0.791
Grade	Low	0	–	–	0	–	–
	High	0.342	0.057	< 0.001	2.007	1	0.157
log(Tumor size)		0.114	0.042	0.008	2.703	1	0.100
Surgery	Not performed	0	–	–	0	–	–
	Performed	-0.945	0.080	< 0.001	14.541	1	< 0.001
Treatment	No therapy	0	–	–	0	–	–
	Chemotherapy	-0.484	0.068	< 0.001	6.817	1	0.009
	Radiotherapy	-0.507	0.124	< 0.001	2.100	1	0.147
	Chemotherapy + Radiotherapy	-0.648	0.057	< 0.001	46.858	1	< 0.001

Table 4: The estimated restricted mean survival times on each group, test statistics comparing the restricted mean survival times of four groups, and p -values when $\tau = 20, 57, 100, 134$ (months)

Treatment	τ			
	20	57	100	134
No therapy	12.330	21.805	28.584	33.125
Chemotherapy	16.125	30.469	41.235	48.555
Radiotherapy	15.005	28.157	40.707	50.012
Chemotherapy + Radiotherapy	16.961	33.278	46.962	56.615
χ^2	233.420	142.474	115.035	101.349
p -value	< 0.001	< 0.001	< 0.001	< 0.001

공변량의 효과를 선형으로 모형화하므로 제한된 평균 생존시간의 회귀모형에서 로그 연결함수를 이용하면 두 모형의 적합 결과를 비교하는 것이 용이하다. Table 5는 $\tau = 20, 57, 100, 134$ 일 때, 로그 연결함수를 사용하여 제한된 평균 생존시간의 회귀모형을 적합한 결과를 보여준다. 모든 τ 에 대해, 진단시 연령이 높을수록 세분화병기와 분화도가 높은 단계일수록, 위암 3기 환자의 평균 생존시간이 감소하였다. 반면 수술과 항암 치료를 받은 경우, 위암 3기 환자의 평균 생존시간이 증가하였다. 이 변수들은 모두 위암 3기 환자의 평균 생존시간에 유의한 영향을 미치는 요인들이었다. 종양의 크기는 연구 초반(20, 57개월)에는 위암 3기 환자의 평균 생존시간을 단축시키는 요인이었으나 연구 후반(100, 134개월)에는 유의한 영향을 미치는 요인은 아니었다. 수술여부 변수의 회귀모수 추정값이 가장 크므로 수술여부가 위암 3기 환자의 평균 생존시간을 늘리는 데 가장 큰 영향을 미치는 요인임을 알 수 있었다. Table 6은 순위-기반 추정법 (Prentice, 1978)을 이용하여 가속화 실패시간 모형을 적합한 결과이다. 유의수준 5%에서 진단연령, 인종, 세분화병기, 분화도, 종양의 크기, 수술여부, 항암치료가 위암 사망 시간에 유의한 영향을 미치는 요인들이었다. 이는 제한된 평균 생존시간의 회귀모형 적합 결과와 같은 결론이다.

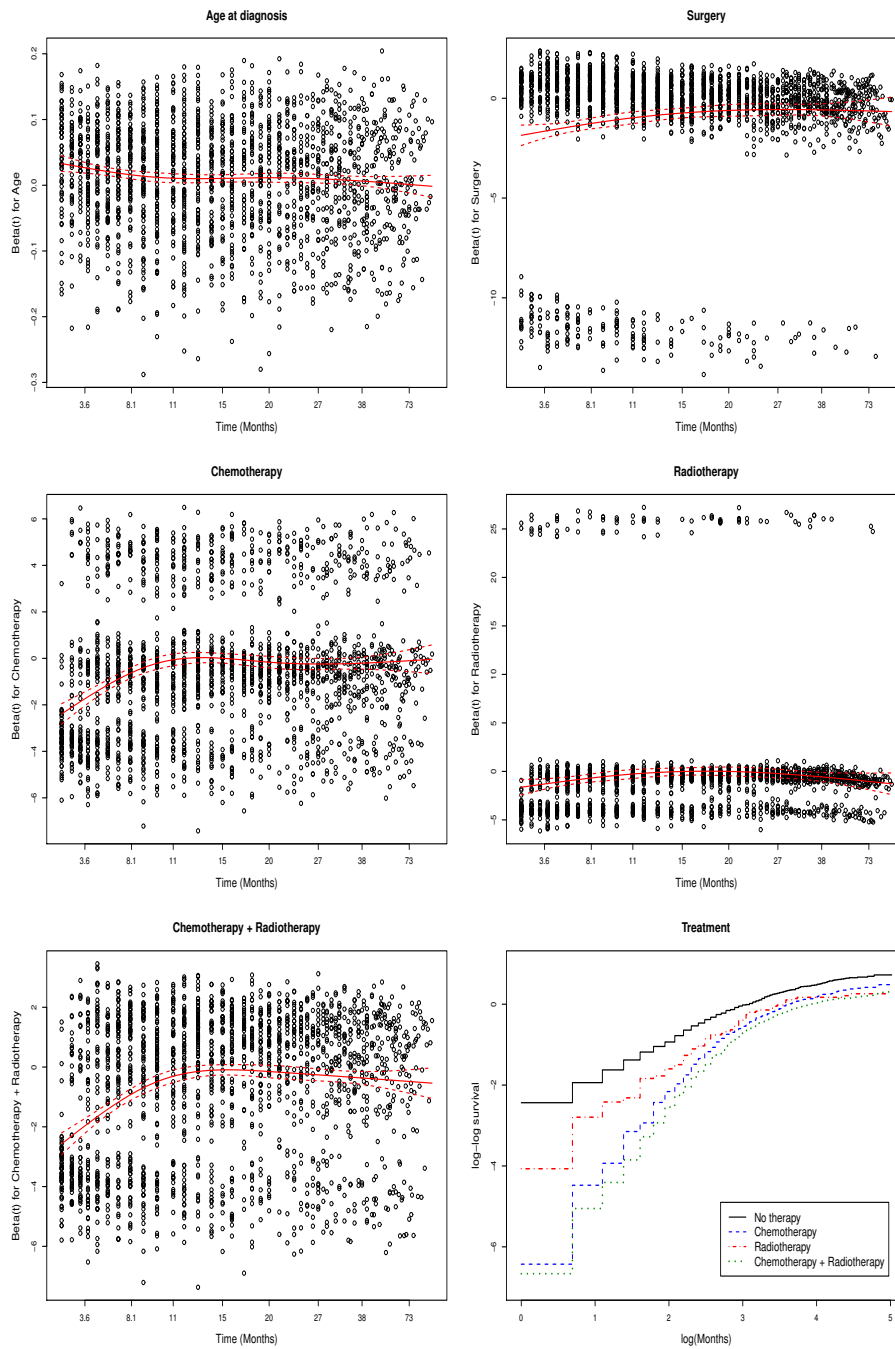


Figure 2: Schoenfeld residual plots for age, surgery, and treatment, and $\log[-\log\{\hat{S}(t)\}]$ plot for treatment.

Table 5: Regression parameter estimates, standard errors, and p-values for death from stomach cancer under the restricted mean survival time model with the log link function when $\tau = 20, 57, 100, 134$ (months)

Covariates		$\tau = 20$			$\tau = 57$		
		$\hat{\beta}$	$se(\hat{\beta})$	p-value	$\hat{\beta}$	$se(\hat{\beta})$	p-value
Intercept		2.597	0.070	< 0.001	3.396	0.137	< 0.001
Age at diagnosis		-0.003	< 0.001	< 0.001	-0.007	0.001	< 0.001
Race	White	0	-	-	0	-	-
	Black	-0.024	0.020	0.231	-0.077	0.042	0.067
	Others	0.057	0.014	< 0.001	0.157	0.029	< 0.001
Substage	IIIA	0	-	-	0	-	-
	IIIB	-0.061	0.016	< 0.001	-0.139	0.032	< 0.001
Grade	Low	0	-	-	0	-	-
	High	-0.082	0.013	< 0.001	-0.201	0.027	< 0.001
log(Tumor size)		-0.027	0.011	0.017	-0.057	0.023	0.013
Surgery	Not performed	0	-	-	0	-	-
	Performed	0.360	0.039	< 0.001	0.600	0.071	< 0.001
Treatment	No therapy	0	-	-	0	-	-
	Chemotherapy	0.226	0.025	< 0.001	0.254	0.045	< 0.001
	Radiotherapy	0.235	0.037	< 0.001	0.283	0.075	< 0.001
	Chemotherapy + Radiotherapy	0.275	0.022	< 0.001	0.332	0.039	< 0.001
Covariates		$\tau = 100$			$\tau = 134$		
		$\hat{\beta}$	$se(\hat{\beta})$	p-value	$\hat{\beta}$	$se(\hat{\beta})$	p-value
Intercept		3.661	0.190	< 0.001	3.798	0.225	< 0.001
Age at diagnosis		-0.008	0.001	< 0.001	-0.009	0.002	< 0.001
Race	White	0	-	-	0	-	-
	Black	-0.084	0.059	0.156	-0.078	0.069	0.262
	Others	0.247	0.040	< 0.001	0.271	0.046	< 0.001
Substage	IIIA	0	-	-	0	-	-
	IIIB	-0.180	0.045	< 0.001	-0.191	0.054	< 0.001
Grade	Low	0	-	-	0	-	-
	High	-0.238	0.038	< 0.001	-0.248	0.045	< 0.001
log(Tumor size)		-0.055	0.032	0.081	-0.054	0.037	0.147
Surgery	Not performed	0	-	-	0	-	-
	Performed	0.703	0.096	< 0.001	0.757	0.112	< 0.001
Treatment	No therapy	0	-	-	0	-	-
	Chemotherapy	0.276	0.061	< 0.001	0.288	0.072	< 0.001
	Radiotherapy	0.374	0.105	< 0.001	0.433	0.124	< 0.001
	Chemotherapy + Radiotherapy	0.390	0.053	< 0.001	0.421	0.063	< 0.001

4. 결론

미국 국립암연구소의 SEER 프로그램으로부터 제공받은 위암 3기 자료에서 항암치료 변수가 비례위험 가정을 만족하지 않아 대안으로 제한된 평균 생존시간을 이용하여 항암치료에 따라 위암 3기 환자의 평균 생존시간에 차이가 있는지 검정하였고, 위암 3기 환자의 평균 생존시간에 유의한 영향을 미치는 공변량들을 파악하기 위해 일반화 선형모형을 통해 회귀모형을 적합하였다. 그 결과, 항암치료법에 따라 위암 3기 환자의 평균 생존시간에는 유의한 차이가 있었으며, 화학요법과 방사선치료 모두 받은 환자의 평균 생존시간이 가장 길었음을 확인하였다. 제한된 평균 생존시간의 회귀모형 적합 결과, 진단시 고령의 나이, 높은 분화도, 세분화 병기 IIIB는 위암 3기 환자의 평균 생존시간을 단축시키는 요인이었고, 수술과 항암치료는 위암 3기 환자의 평균 생존시간을 늘리는 요인이었다. 종양의 크기가 클수록 연구 초반에는 위암 3기 환자의 평균 생존시간을 단축시켰으나 연구 후반에는 통계적으로 유의한 영향을 미치는 요인이 아니었음을 확인하였다. 수술여부가 위암 3기 환자의 평균 생존시간을 늘리는데 가장 큰 영향을 미치는 요인이었다. 항암치료가 수술 후 환자의 상태를 호전시키는 역할을 할 뿐만 아니라 수술 전 환자의 상태를 수술이 가능하게 함으로써 수술을 통해

Table 6: Regression parameter estimates, standard errors, and p -values for death from stomach cancer under the accelerated failure time model

Accelerated failure time model		$\hat{\beta}$	$se(\hat{\beta})$	p -value
Age at diagnosis		-0.020	0.002	< 0.001
Race	White	0	-	-
	Black	-0.147	0.047	0.002
	Others	0.493	0.078	< 0.001
Substage	IIIA	0	-	-
	IIIB	-0.354	0.079	< 0.001
Grade	Low	0	-	-
	High	-0.480	0.074	< 0.001
log(Tumor size)		-0.132	0.058	0.024
Surgery	Not performed	0	-	-
	Performed	1.218	0.140	< 0.001
Treatment	No therapy	0	-	-
	Chemotherapy	0.665	0.083	< 0.001
	Radiotherapy	0.833	0.229	< 0.001
	Chemotherapy + Radiotherapy	0.907	0.073	< 0.001

미세 종양을 제거할 수 있고 이는 위암 3기 환자의 평균 생존시간을 늘리는데 도움을 준다. 이런 사실 때문에 항암치료에 관한 많은 연구가 수술과 항암치료의 순서에 대한 변수를 포함하고 있지만 분석에 사용된 SEER 위암 3기 자료는 이런 정보를 가지고 있지 않다. 본 논문의 분석에서 수술과 항암치료의 순서에 관한 변수를 포함하면 항암치료 효과를 좀 더 세분화하여 비교함으로써 더 유의미한 결론을 얻을 수 있을 것이라 생각된다. 이는 본 연구의 향후 과제가 될 것이다. 식습관에 관한 정보가 제공된다면 이를 공변량으로 고려하여 위암 환자의 평균 생존시간에 미치는 영향을 조사해 보는 것도 의미있을 것이다. 이는 본 연구의 향후 연구과제가 될 것이다. 본 논문에서 사용되어진 제한된 평균 생존시간에 관한 분석방법들은 다른 암 자료의 분석에도 활용가능하며 암 환자의 평균 생존시간을 늘리기 위한 정보를 얻는데 도움이 될 것이라 기대한다.

References

- Andersen PK, Hansen MG, and Klein JP (2004). Regression analysis of restricted mean survival time based on pseudo-observations, *Lifetime Data Analysis*, **10**, 335–350.
- Andersen PK and Perme MP (2010). Pseudo-observations in survival analysis, *Statistical Methods in Medical Research*, **19**, 71–99.
- Chen PY and Tsiatis AA (2001). Causal inference on the difference of the restricted mean lifetime between two groups, *Biometrics*, **57**, 1030–1038.
- Cox DR (1972). Regression models and life-tables (with discussion), *Journal of Royal Statistical Society, Series A*, **34**, 187–220.
- Grambsch P and Therneau T (1994). Proportional hazards tests and diagnostics based on weighted residuals, *Biometrika*, **81**, 515–526.
- Guo C and Liang Y (2019). Analyzing restricted mean survival time using SAS/STAT, *SAS Institute Inc., Cary, NC*.
- Irwin JO (1949). The standard error of an estimate of expectation of life, with special reference to expectation of tumourless life in experiments with mice, *The Journal of Hygiene*, **47**, 188–189.

- Kaplan EL and Meier P (1958). Nonparametric estimation from incomplete observations, *Journal of the American Statistical Association*, **53**, 457–481.
- Korea Central Cancer Registry (2019). National cancer center, Annual report of cancer statistics in Korea in 2017, Ministry of health and welfare.
- Liang KY and Zeger SL (1986). Longitudinal data analysis using generalized linear models, *Biometrika*, **73**, 13–22.
- National Cancer Information Center (2019). Cancer information service. from: <https://www.cancer.gov/>
- Nemes S, Bülow E, and Gustavsson A (2020). A brief overview of restricted mean survival time estimators and associated variances, *MDPI Stats*, **3**, 107–119.
- Prentice RL (1978). Linear rank tests with right censored data, *Biometrika*, **65**, 167–180.
- Royston P and Parmar MK (2011). The use of restricted mean survival time to estimate the treatment effect in randomized clinical trials when the proportional hazards assumption is in doubt, *Statistics in Medicine*, **30**, 2409–2421.
- Ruhl J, Ward E, Hofferkamp J, *et al* (2019). Grade Manual. *NAACCR*, Springfield, IL 62704–4194.
- Schoenfeld D (1982). Partial residuals for the proportional hazards regression model, *Biometrika*, **69**, 239–241.
- Surveillance, Epidemiology, and End results (SEER) program (www.seer.cancer.gov) SEER*Stat Database : Incidence - SEER 18 Regs custom data (with additional treatment fields), Nov 2018 Sub (1975-2016 varying) - Linked to county attributes - total US, 1969-2017 counties, national cancer institute, DCCPS, Surveillance research program, released April 2019, based on the November 2018 submission.
- Tian L, Zhao L, and Wei LJ (2014). Predicting the Restricted mean event time with the subject's baseline covariates in survival analysis, *Biostatistics*, **15**, 222–233.
- Zeger SL and Liang KY (1986). Longitudinal data analysis for discrete and continuous outcomes, *Biometrics*, **42**, 121–130.
- Zhao L, Tian L, Uno H, Solomon SD, Pfeffer MA, Schindler JS, and Wei LJ (2012). Utilizing the integrated difference of two survival functions to quantify the treatment contrast for designing, monitoring, and analyzing a comparative clinical study, *Clinical Trials*, **9**, 570–577.

Received December 22, 2020; Revised January 27, 2021; Accepted February 22, 2021

제한된 평균 생존시간을 이용한 위암 3기 자료 분석에 관한 연구

김빛나^a, 이민정^{1,a}

^a강원대학교 통계학과

요 약

본 연구는 미국 국립암연구소의 SEER 프로그램에서 제공하는 위암 3기 자료에 대해 항암치료의 효과를 비교하고 위암 생존율에 유의한 영향을 미치는 요인을 알아보고자 한다. 본 연구에서 분석한 위암 3기 자료는 비례위험 가정이 성립하지 않아 대안으로 제한된 평균 생존시간을 이용한 분석 방법을 자료 분석에 적용하였다. 의사-관측들을 이용하여 제한된 평균 생존시간을 추정하였고, 제한된 평균 생존시간 추정량에 기반한 검정통계량을 이용하여 항암치료의 효과를 파악하였다. 일반화 선형모형을 이용한 회귀모형을 통해 위암 3기 환자의 평균 생존시간에 유의한 영향을 미치는 공변량들의 효과를 추정하였다. 항암치료법에 따라 위암 3기 환자의 평균 생존시간에 유의한 차이가 있음을 확인하였고, 진단연령, 인종, 세분화병기, 분화도, 종양의 크기, 수술여부, 항암치료가 위암 3기 환자의 평균 생존시간에 유의한 영향을 미치는 요인들이었으며, 그 중 수술여부가 위암 3기 환자의 평균 생존시간을 늘리는데 가장 큰 영향을 미치는 요인임을 확인하였다.

주요용어: 생존분석, 제한된 평균 생존시간, 의사-관측, 위암 자료

이 논문은 2018년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임 (NRF-2018R1D1A1 B07041070). 본 연구는 제1저자 김빛나의 강원대학교 석사학위논문 일부 발췌, 수정한 논문임.

¹교신저자: (24341) 강원도 춘천시 강원대학길 1, 강원대학교 통계학과. E-mail: mlee@kangwon.ac.kr